UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

POU1F1 mutations and hypopituitarism: Relationship with the clinical phenotype of hypopituitarism, molecular modeling and functional prediction.

Verónica Lucía Castañeda Pérez

Ingeniería en Procesos Biotecnológicos

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniera en Procesos Biotecnológicos

Quito, 4 de mayo de 2020

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN
DE TRABAJO DE INTEGRACIÓN CURRICULAR

POU1F1 mutations and hypopituitarism: relationship with the clinical phenotype of hypopituitarism, molecular modeling and functional prediction

Verónica Lucía Castañeda Pérez

Nombre del profesor, Título académico **Andrés Caicedo, Ph.D.**

Quito, 4 de mayo de 2020

DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Nombres y apellidos:	Verónica Lucía Castañeda Pérez
Código:	00132267
Cédula de identidad:	1723309496
Lugar y fecha:	Quito 4 de Mayo de 2020

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

La Deficiencia Combinada de Hormonas de la Pituitaria tipo 1 (CPHD1), también conocida como panhipopituitarismo, es una enfermedad causada por mutaciones en el factor de transcripción POU1F1. Estos cambios en la secuencia del gen afectan negativamente la capacidad de POU1F1 para inducir la producción de hormonas desde la hipófisis anterior, lo que impide el desarrollo. Diferentes mutaciones en el gen POU1F1 (Pit-1, GHF-1) determinan la severidad de CPHD1. Estudiar las mutaciones en POU1F1 ha llevado a importantes avances en la comprensión de CPHD1. Sin embargo, no se ha aclarado completamente cómo cada mutación afecta la funcionalidad de la proteína y cómo se relacionan con un fenotipo particular, lo que perjudica el desarrollo de terapias dirigidas a cada caso. Por lo tanto, el objetivo de este estudio es proponer el uso de bioinformática como una herramienta para asociar los cambios funcionales debido a mutaciones POU1F1 con el fenotipo clínico. Se realizó una revisión sistemática de la literatura (RSL) de Pit-1-β (de 291 aminoácidos) para encontrar las mutaciones descritas en trabajos previos. Se encontraron un total de 45 mutaciones patológicas, y se estudiaron a profundidad las 24 mutaciones sin sentido. Las mutaciones en el exón 4 están relacionadas principalmente con pacientes europeos, mientras que las del exón 6 son más comunes en pacientes árabes y asiáticos. Se clasificaron las mutaciones sin sentido usando filogenia y así se encontraron cuatro grupos de mutaciones. Pit-1-β se modeló utilizando I-TASSER, incluidas las regiones intrínsecamente desordenadas: TAD (dominio de transactivación) y la región de enlace para estudiar secuencias mutadas. Con base en la estructura original, modelamos las proteínas mutantes y se analizó su unión a CBP y ADN a través del acoplamiento molecular. Estos hallazgos proporcionan una mejor comprensión de CPHD1 y la posibilidad de desarrollar tratamientos personalizados para el paciente.

Palabras clave: Deficiencia combinada de hormona pituitaria, factor de transcripción, plegamiento de proteínas, revisión sistemática de la literatura

ABSTRACT

Combined Pituitary Hormone Deficiency type 1 (CPHD1), also known as panhypopituitarism, is a disease caused by mutations in the transcription factor POU1F1. These gene sequence changes negatively affect the ability of POU1F1 to induce anterior pituitary hormone production, hindering development. Different mutations in the POU1F1 gene (Pit-1, GHF-1) determine the severity of CPHD1. Studying mutations in POU1F1 has led to important advances in understanding CPHD1. However, it is not fully understood how each mutation affects the functionality of the protein and how they are related to a particular phenotype, which hinders the development of therapies directed at each case. Therefore, the objective of this study is to propose the use of bioinformatics as a tool to associate functional changes due to POU1F1 mutations with the clinical phenotype. A systematic literature review (RSL) of Pit-1-β was performed to find the mutations described in previous work. A total of 45 pathological mutations were found, and the 24 nonsense mutations were studied in depth. Mutations in exon 4 are mainly related to European patients, while those in exon 6 are more common in Arab and Asian patients. Missense mutations were classified using phylogeny and we found four groups of mutations. Pit-1-β was modeled using I-TASSER, including the intrinsically disordered regions: TAD (transactivation domain) and the binding region to study mutated sequences. Based on the original structure, we modeled the mutant proteins and analyzed their binding to CBP and DNA through molecular coupling. These findings provide a better understanding of CPHD1 and the possibility of developing personalized treatments for every patient.

Key words: Combined pituitary hormone deficiency, transcription factor, protein folding, systematic literature review

TABLE OF CONTENTS

TABLES INDEX

FIGURES INDEX

1. INTRODUCTION

Mutations hamper POU1F1's capacity to support development of the anterior pituitary gland and activate the production of Growth Hormone (GH), Prolactin (PRL) and Thyroid-Stimulating Hormone (TSH) (Radovick et al., 1992). CPHD cases may present with short stature and/or hypothyroidism and may also be associated with mental retardation (Radovick et al., 1992). The clinical features may include a prominent forehead, marked midfacial hypoplasia with depressed nasal bridge, deep-set eyes, and a short nose with anteverted nostrils, along with other midline abnormalities and a hypoplastic pituitary gland determined by magnetic resonance imaging (MRI) (Aarskog et al., 1997). Different mutations in the POU1F1 (Pit-1) gene have been associated with a variety of phenotypes and severity of CPHD1, hence, understanding these variations would be important to provide personalized treatment to the patient.

The focus of this systematic literature review (SLR) is to provide a detailed list of the mutations in the POU1F1 gene, splice variant β, consisting of 291 aminoacids. As a result, the most common variants and associations with specific populations were determined. This research aims to identify the most severe mutations of POU1F1 and explain the phenotype through structural differences. The naïve protein structure of POU1F1 was obtained guided by bioinformatics, to allow future topological analysis. The naïve structure includes the Transactivation Domain (TAD) and linker region, both of which have not been previously obtained due to their flexibility. The objective, using the SLR and molecular meta-analysis, is to provide a better understanding of the origin of CPHD1 and provide information for the development of innovative and personalized therapeutic approaches.

1.1. CPHD symptoms, characterization and treatment

The key clinical findings of CPHD include short stature with growth hormone (GH) deficiency, central hypothyroidism with a low or inappropriately normal TSH, and a hypoplastic pituitary gland. Some patients also present with various degrees of developmental abnormalities which may be due to hypoglycemia during the neonatal period due to severe GH deficiency (Fang et al., 2016). CPHD is classified in CPHD1, CPHD2, CPHD3, CPHD4 and CPHD6 based on the mutations of key genes associated with hormone deficiency (Baş et al., 2015a), respectively POU1F1, PROP1, LHX3, LHX4 and OTX2 (*Combined pituitary hormone deficiency - Genetics Home Reference - NIH*, n.d.).

CPHD is treated with hormonal replacement after the clinical diagnosis is made. Delay in diagnosis results in further accentuation of abnormalities associated with growth and development. Treatment of CPHD requires lifelong hormonal replacement. Earlier treatment with GH or thyroid hormone improves patients' prognosis and development (Sadeghi-Nejad & Senior, 1974).

1.2. Molecular characteristics of POU1F1 and its mutations:

Missense, nonsense and splicing mutations of POU1F1 may cause CPHD1 (Kinoshita et al., 1994). CPHD1 can be sporadic or hereditary (Fang et al., 2016). Cases of families with a dominant or a recessive pattern of inheritance have been reported (*Combined pituitary hormone deficiency - Genetics Home Reference - NIH*, n.d.). 90% of patients diagnosed with CPHD had a non-hereditary cause and therefore of unknown etiology (De Rienzo et al., 2015).

POU1F1 belongs to the POU family of mammalian transcription factors with well conserved homeodomains. POU1F1 has two DNA binding regions, the POU-specific and the POU-homeodomains, forming a homodimer dimer on DNA. Additionally, similar proteins have been reported in other animals, such as: *Caenoharbditis elegans* (Herr et al., 1988a), *Xenopus laevis* (Cao et al., 2007), *Drosophila melanogaster* (Johnson & Hirsh, 1990), turkeys (Bastos et al., 2006), chickens (Bastos et al., 2006), jellyfish (Hroudova et al., 2012), and various fish species (Bastos et al., 2006). POU1F1 paralogues include: Pit-1, Oct-1, Oct-2 and Unc-86 (Karp, 2009), they are transcription factors as well and are responsible for supporting organism development especially the nervous system.

The consensus sequence to which POU1F1 binds is 5'-TAAAT-3' (*POU1F1 - Pituitary-specific positive transcription factor 1 - Homo sapiens (Human) - POU1F1 gene & protein*, n.d.). Pit-1 has been synthesized and expressed for crystallography; Pit-1_{wt} included residues 1 to 291, Pit-$1_{\Delta N}$ residues 105–291 (portraying the POU_S and POU_{HD}) and Pit-$1_{\Delta N \Delta C}$ residues 124–273 (POU_{HD} only) (Agarwal & Cho, 2018). Results of its crystallographic image describe the binding of the double helix to a 4xCATT palindromic repeat and the formation of a total of eight helixes. In humans, five of these helixes are formed in the POU_S and 3 of them in the POU_{HD} (Agarwal & Cho, 2018).

The inner side of the helixes is positively charged and hydrophilic, while the outer part hydrophobic allowing its movement into the nucleus (*POU1F1 - Pituitary-specific positive transcription factor 1 - Homo sapiens (Human) - POU1F1 gene & protein*, n.d.). The DNA binding helixes, POU_S and POU_{HD}, are connected by a 15 amino acid linker region, providing the flexibility

needed to bind to the DNA sequence found on different promoter regions. Another important region of this protein is the TAD which is a transcription factor scaffold domain which is generally acidic and rich in glutamine, proline and isoleucine (Wärnmark et al., 2003). POU1F1 has a phosphorylated region as a consequence of its interaction with CBP/p300 (Creb-Binding Protein), which allows the activation and functionality of POU1F1 (R. N. Cohen et al., 2006).

1.3. Bioinformatics

Protein modelling allows its visualization and understanding of its functionality. Nowadays, there are many tools used to obtain a protein model and structure. For example, the I-TASSER tool uses reported structures and recreates new proteins using its homology with known sequences, additionally it also creates them *ab initio* (J. Yang & Zhang, 2015). The *ab initio* modelling refers to modelling the structure using physical principles rather than the homology. The advantage of using I-TASSER is the Residue-Specific Quality (RSQ) component which allows to obtain more efficient models avoiding errors. The estimation of the of protein structure prediction is made by the RSQ, the B-factor profile (BFP) analyze all residues along the chain, the C-score and the confidence level provides information on each atom's position (J. Yang et al., 2016). The C-score provides a confident score, considered appropriate between -5 and 2, and therefore we use it to determine which of the obtained models is the most reliable (J. Yang & Zhang, 2015).

Another strategy to evaluate how affected is the protein by specific mutations is by using its wildtype or normal model and recreate the mutation. This allows the visualization of the rotamers

and the different positions that the mutation could induce in the protein. With this strategy it is possible to identify if a particular mutation clash with the structure and can be done using the program Chimera from UCSF (Z. Yang et al., 2012).

A limitation of using I-TASSER modelling is that some proteins lack a fixed 3D structure. These are often referred to as intrinsically disordered proteins. In this case, POU1F1, as other transcription factors, has flexible domains. Meaning that some domains in POU1F1 can change their structure when interacting with other proteins and with the DNA. However, Intrinsically Disordered Protein Databases can be used to predict the regions where loops and similar structures are more likely to form and it is possible to compare this result to the result from the modelling (Linding et al., 2003a).

The mutant modelled protein can have a significantly different structure than the *wt*. However other changes can be induced by missense mutations such as the physical and chemical properties which affect the protein's functionality. Molecular docking can be used to evaluate different protein properties (Cheng et al., 2007). Therefore, with this approach is possible to we can study the ability of POU1F1 to interact with CBP and to DNA using molecular docking.

2. METHODOLOGY

2.1 Systematic Literature Review and data extraction

The systematic literature review was the basis for the research. Through it the study gains broad perspective about CPHD. Search terms were defined for articles to retrieve information from SCOPUS as follows (Annex A contains more detail about search terms and Annex B about the article classification):

((TITLE-ABS-KEY (*POU1F1*)) OR (TITLE-ABS-KEY (GHF-1)) OR (TITLE-ABS-KEY (*Pit-1*)))AND ((TITLE-ABS-KEY (human*)) OR (TITLE-ABS-KEY (homo sapiens)) OR (TITLE-ABS-KEY (person*)) OR (TITLE-ABS-KEY (people))) AND ((TITLE-ABS-KEY (structure)) OR (TITLE-ABS-KEY (aminoacid*)) OR (TITLE-ABS-KEY (peptid*)) OR (TITLE-ABS-KEY (Protein AND structure)) OR (TITLE-ABS-KEY (*sequence*)) OR (TITLE-ABS-KEY (proteom*)) OR (TITLE-ABS-KEY (transcript*)) OR (TITLE-ABS-KEY (exome*)) OR (TITLE-ABS-KEY (exon*)) OR (TITLE-ABS-KEY (evolution)) OR(TITLE-ABS-KEY (mut*)) OR (TITLE-ABS-KEY (homolog*)) OR (TITLE-ABS-KEY (paralog*)) OR (TITLE-ABS-KEY (gen*)) OR (TITLE-ABS-KEY (molecular AND modeling)) OR (TITLE-ABS-KEY (splicing)) OR (TITLE-ABS-KEY (functional AND prediction)) OR (TITLE-ABS-KEY (DNA AND binding))

Later, the articles were screened to extract relevant data for analysis, especially the ones referring to the structure and clinical phenotype (classification in Annex B).

Missense mutations show only one amino acid change, and therefore do not radically change the protein's structure. Acknowledging that, the structural, pathological and demographic inferences were based on the 24 missense mutations. For the pathological analysis, the data was summarized in a pathogenicity matrix. A matrix counting was created, with one occurrence per patient and clustering those phenotypes by systems, taking into account developmental process and a molecular etiology approach (see classification in Annex C). In addition, novel mutations that cause alternative splicing and premature stop codons are described in SLR results. The symptoms consistent with CPHD were described and analyzed for each mutation. If one mutation

is reported more than once, then all the symptoms will be described and used for following analysis, so that the full spectrum is represented.

2.2 Protein reconstruction and evo-devo-med analysis

For the wild type (*wt*), the amino acid sequence was retrieved from Uniprot (P28069), using the canonical Pit-1-β isoform. Mutants were manually created in MEGA X and sequences visualized by the use of ESPript 3.0 (Robert & Gouet, 2014). Swiss model server was first used to predict an initial structure whit linker region between POUS and POUHD (Bertoni et al., 2017; Guex et al., 2009; Jasper J. Koehorst, Jesse C. J. van Dam, Edoardo Saccenti, Vitor A. P. Martins dos Santos, 2017). This versatile software allows the prediction of the previously unobtained linker region from the initial restrained PDB (5wc9) (Agarwal & Cho, 2018). Next, this structure was used as a template to obtain the final *wt* isoform within the TAD region. In this way, the use of two different servers allows the prediction of a final model within previously unreported regions. This high-resolution structure was used as a template for all mutant models obtained using I-TASSER server (J. Yang & Zhang, 2015; Zhang et al., 2017).

To choose the best model, a DNA was performed to match with the corresponding chains of the reference PDB (5wc9) model (Agarwal & Cho, 2018) and dock with the CBP using the PyDock server. From the five resulting models, only model 4 had best fit with DNA. The structure was also verified using DisEMBL (Linding et al., 2003b) and PONDR® (Dunker et al., 2001; Garner et al., 1999) VLXT, and VSL2 algorithms to analyze the intrinsically disordered amino acids of POU1F1's structure, described in the DisProt database (Hatos et al., 2020; Piovesan et al., 2017).

Mutant proteins were obtained using two different methods. The first one using I-TASSER and the second one with Chimera. First, the model 4 wildtype structure of POU1F1 acted as a template to model the mutant proteins through I-TASSER. Secondly, using the wildtype structure for POU1F1, we generated each mutation using Chimera. This simulation allows the visualization of the possible rotamers for every mutant aminoacid. There was the possibility that the mutations caused a clash or contacts within the internal structure, which would affect the protein's structure and could also be used to explain the phenotype (Pettersen et al., 2004).

To analyze the 24 missense mutants and associate them with their pathological data, first the summarized SLR results from pathological matrix were presented into a heat map chart. Then we constructed a dendrogram based on model structure predictions. Tree was predicted by the use of the DALI server (Holm, 2019), and visualized with Fig Tree software (Rambaut & FigTree, 2018).

3. RESULTS

3.1 Systematic Literature Review

As a result of the SLR we found 1 194 articles. The clinical reports presented various types of mutations that cause CPHD. The following types of mutations were described previously in the public HGMD: missense (20), alternative splicing (4), small deletions (3), small insertions (3), gross deletions (2) and regulatory mutations (2), in total 34 mutations (Stenson et al., 2014). However, 1 new nonsense and 2 new missense mutations were found through the SLR of clinical articles. These are: R143X (Blum et al., 2018a), Q167R (Gavrilova et al., 2017), and I244S (Baş et al., 2015a). Four novel mutations in the promotor region were also found in the SLR, including

rs300996 (Sun et al., 2006), − 1295C > T (Elizabeth et al., 2018), rs10511134 (T > A) (Elizabeth et al., 2018), and rs300982 (C > T) (Elizabeth et al., 2018). One new mutation that affect protein splicing were also found in the SRL: c.605-1G > A. Three other novel mutations found were c.605delC (Birla et al., 2016), c.1-59T > A (Birla et al., 2016) , and c. + 8C > T (Birla et al., 2016). Therefore, 45 total mutations were reported in humans. Other mutations have been studied since they were found in other species, such as Q4R (Pernasetti et al., 1993) and D227Y (Pernasetti et al., 1993), or simply reported as a variant like A19V (*UniProtKB/SwissProt variant VAR_049361*, n.d.). Other mutations have been found in healthy individuals such as R113W (De Graaff et al., 2010) and G89R (Reynaud et al., 2006) which are therefore considered polymorphisms. From the 25 missense mutations, 8 were dominant (32%) and 17 recessives (68%) (every mutation and ethnicity are presented in Annex D).

Missense mutations were first classified according to their corresponding domain, then related to the ethnicity of the patient that presented with the mutation (Figure 1). Most mutations were reported only once or twice, except for the mutation R271W which has been the most reported in this SLR. All the cases presented in the NCBI Clinvar database also included in this study.

To analyze the pathological phenotype, the mutations were mapped according to their domain and also per mutation (Figure 2). CPHD can be diagnosed at birth when it presents with growth hormone deficiency and hypoglycemia, during newborn screening due to a low thyroid hormone levels and low or normal TSH. However, some of the described cases were diagnosed later in life, which may affect symptoms identification and correct interpretation. Most patients' symptoms are caused by a deficiency in GH, TSH and PRL, which is why growth is the most

affected characteristic of CPHD patients. Some infants may have facial bone structure is also affected, such as a prominent forehead, and depressed nose bridge (Pellegrini-Bouiller et al., 2008). Later in life, children present with short stature and failure to thrive and may at that time be noted to have facial characteristics (Lee et al., 2011). It is critical to diagnose children with CPHD early as treatment with GH replacement and thyroid hormone is highly effective and most beneficial when begun early (Fang et al., 2016).

Once CPHD is suspected, a molecular analysis is carried out using PCR and sequencing (Fang et al., 2016). Only one case of a mutation in POU1F1 has been described in Latin America, this was R271W in Brazil (Taha et al., 2005), no cases have been described in Central and South African countries. Whether this is due to a lack of testing is not clear from the literature.

As a result of the SLR, structural data was collected to characterize the functional parts of the protein represented in the figures. Four main parts of the protein were identified accordingly to the literature, the Transactivation domain (TAD) the POU specific domain (POU_S) the linker region (Linker) and the POU homeodomain (POU_{HD}).

3.1 Protein modelling

Five models were obtained from I-TASSER server, with TM-scores that range from -4.59 to -3.35. Using these methods, the best model is structure number 1, however, structure #4 (model 4) was chosen as the most adequate since Pydock analysis with the DNA sequence from the 5wc9 model and binding with CBP presented the least problematic topology. These differences are denoted by the quantum performance folding of each structural homolog. The TAD is composed

of three small helixes and many coiled coil loops. The first one from amino acid (aa) 32 to 34 (AEC), the next one from aa 38 to 40 (SNH) and the third 68 to 71 (GVMA) and is represented in Figure 3. POU1F1 is phosphorylated by protein kinase A (PKA) on three amino acids; Serine 119, Glycine 220, and Threonine 220 (Agarwal & Cho, 2018).

Additionally, other sites were used to analyze the protein's intrinsically disordered regions. First, the blast of POU1F1's aminoacid sequence in DisProt and obtained 12 results. Among those, the one with the lowest E-value was 3e-50, with 0.29 Identity and 0.59 coverage. The most similar protein presented in this database was POU2F1 (MobiDB:P14859, UniProtKB:P14859) and presented 30.3% of the disordered content. Two regions in POU2F1 showed a disordered structural state, the first one a fragment before the POUS and another one between the POU domains. The second server was DisEMBL which predicted three regions to be disordered. This server uses three criteria: the formation of loops/coils, hot loops, and missing coordinates (Remarks-465). The next server used was PONDR® to find the disordered regions in POU1F1, using the VLXT, and VSL2 algorithms. Results are summarized in Table 1, noting that the I-TASSER model excludes helixes corresponding to the POUS and POUHD.

Figure 4 shows the modelled protein with the entire 291 amino acid sequence compared to the crystallographic structure which only represents the POU domains. The non-aligned region corresponds to the TAD and linker. The alignment was performed with TM-align server where the reference protein is the crystallographic structure.

3.2 Pathologic structures reconstruction

Reconstructed proteins uploaded to the TM-align present values above 0.98 in every case except for R143Q (0.51) The alignment is rated from 0 to 1 where 1 is an identical match, therefore the mutant protein structures (obtained through I-TASSER) were almost identical (Annex F, protein models are shown in Annex G).

Data from the pathogenic matrix of the 24 mutants is presented as a heat plot showing the most pathogenic phenotypic diverse mutants per system (For the complete phenotypic tabulation see Annex E). The resulted model showed to be predictive of the thyroid pathology (TH) and development and growth delay (DG). Following TH and DG, Growth Hormone Deficiencies (GH) is the second most pathogenic predicting. The F233S mutation using the DG classification, is predicted to be the most pathogenic. The remainder of the CPDH mutations present a medium or low prediction for pathogenicity (Figure 5). For example, DALI server predicts a tree with a well-established topology. A basal z-score support of 14.5 divides the dendrogram from the root into two large clusters that present long branches after cluster expansion. The first one by the R143Q mutation that performs as an initial branching and a second group, that is more numerous and consist of three subgroups (See Annex H for z-score matrix). As result, the tree presents four main clusters, grouped by their folding relationships and sequence homologies (Figure 5). In ascending order, the groups are color coded as orange, yellow, green and blue.

4. TABLES

Table 1. Loops predicted for the disordered segments

Criteria	***Amino acid number***
Loops/coils	1-105, 117-128, 160-168, 206-218, 231-240
Hot loops	1-17, 113-130, 202-220, 225-241, 264-291
I-TASSER	3-8, 32-34, 38-40, 68-71, 85-88, 115-118, 203-206, 272-275, 281-289

5. FIGURES

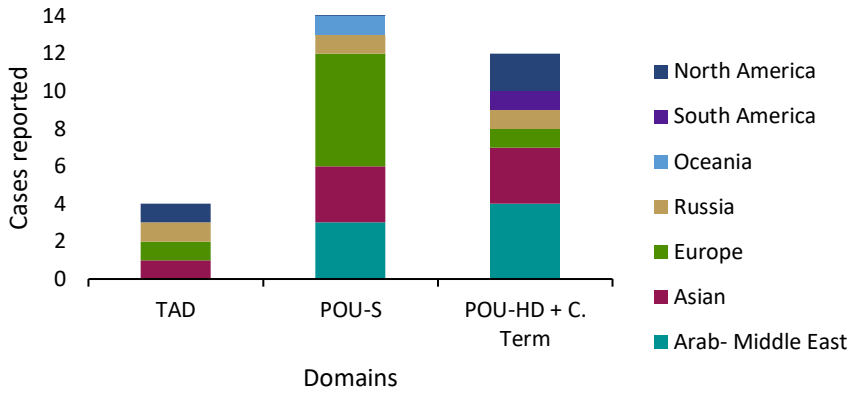


Figure 1. Reported missense mutations as clinical cases.

These 24 missense mutations represent also the mutations analyzed for the clinical phenotype and structural analysis. **Figure 1** classifies the number of cases as well, organized according to its corresponding domain. Exons 1, 2 and part of exon 3 correspond to the TAD. The main difference between them is that exon 3 corresponds to both, TAD and POUs. The POUs contains exons 3 and 4. The mutations in exon 5 and 6 are included in the POUHD. We analyze 31 cases in total. Only one missense mutation was present in exon 5, the K216E mutation in a North American patient. The R271W mutation was present in Asia, Europe, North and South America. The only mutation presented in the C-terminal (C. Term, part of exon 6) corresponds to the R271W mutation.

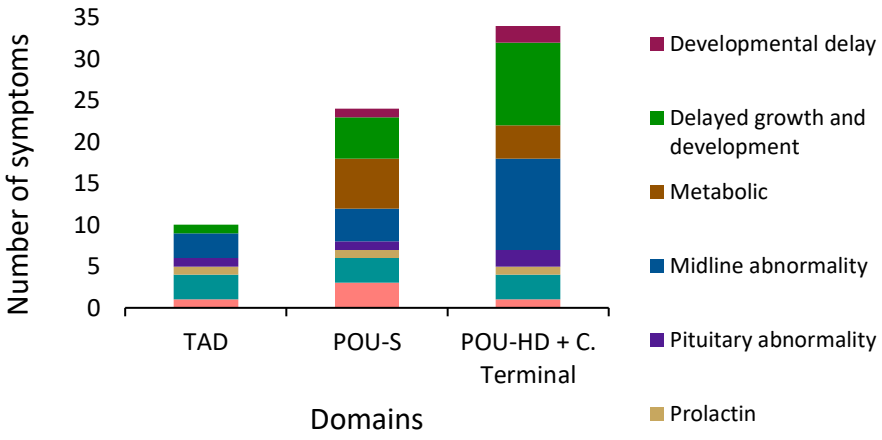


Figure 2A

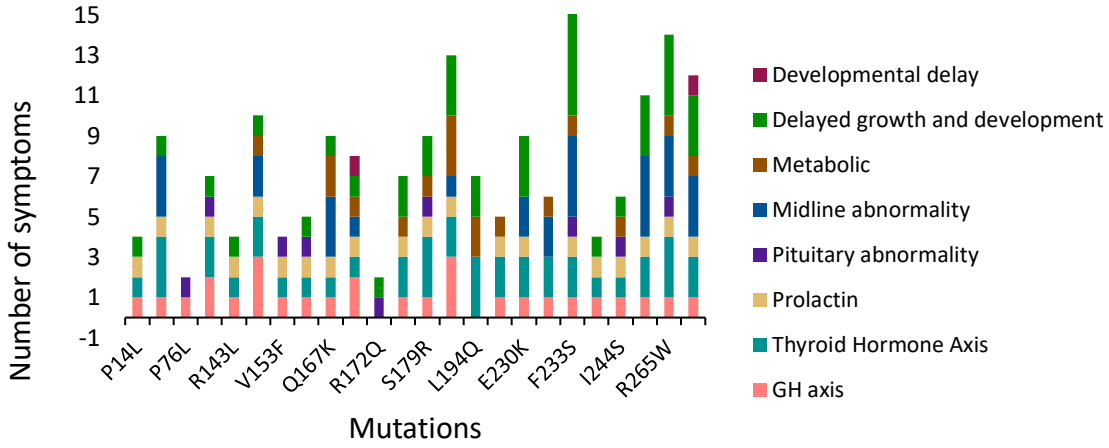
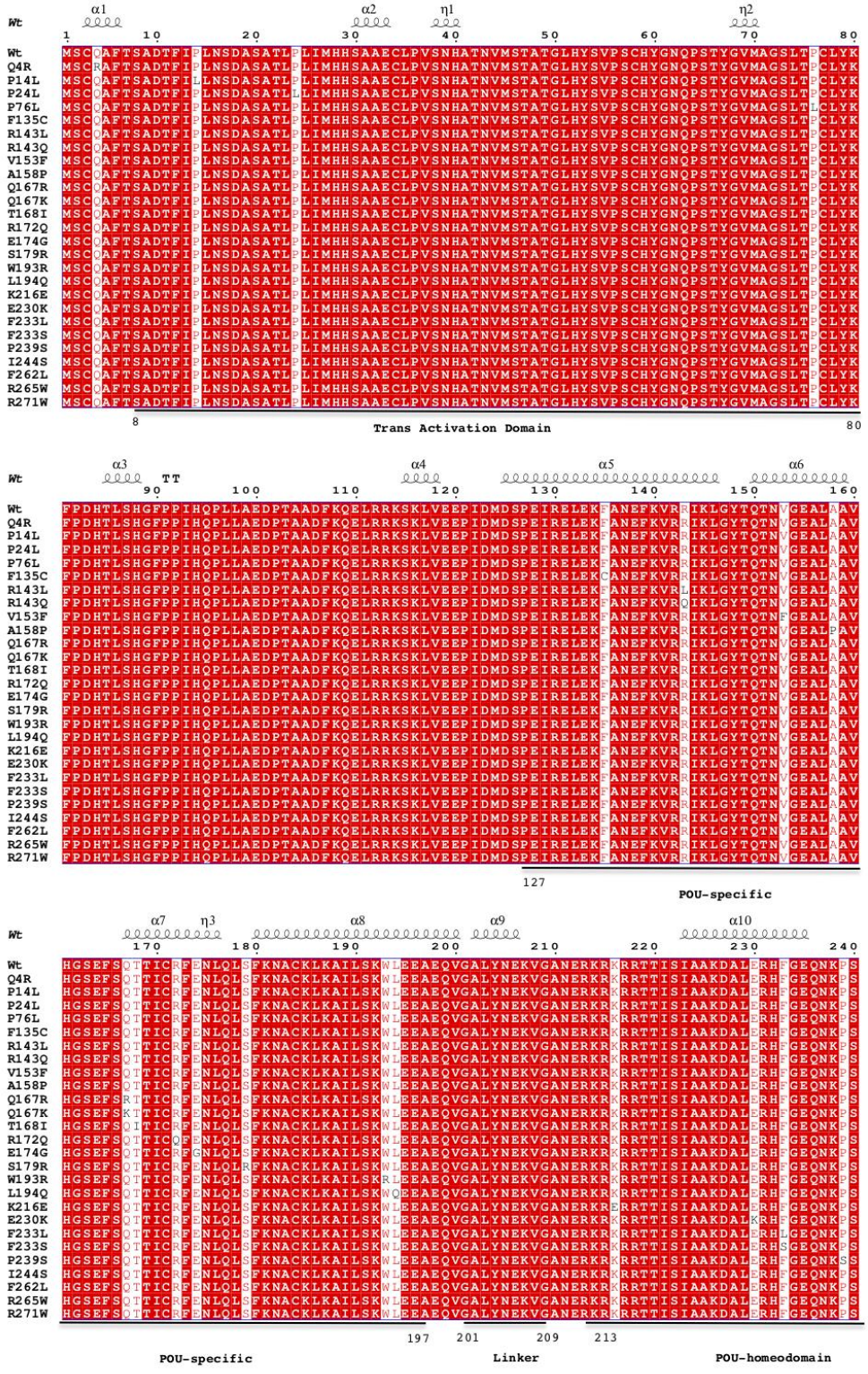


Figure 2B

Figure 2. Clinical phenotypes summarized per system affected.

Figure 2 was constructed based on development and system pathology approach. The symptoms consisting with the pathology were classified according to the table in Annex C, all symptoms are presented in Annex E. In Figure 2A each mutation was associated with its corresponding domain, so it compares the diversity of pathologies by affected systems per domain. Figure 2B presents each system affected by each mutation individually.


Trans Activation Domain
POU-specific
POU-specific
Linker
POU-homeodomain

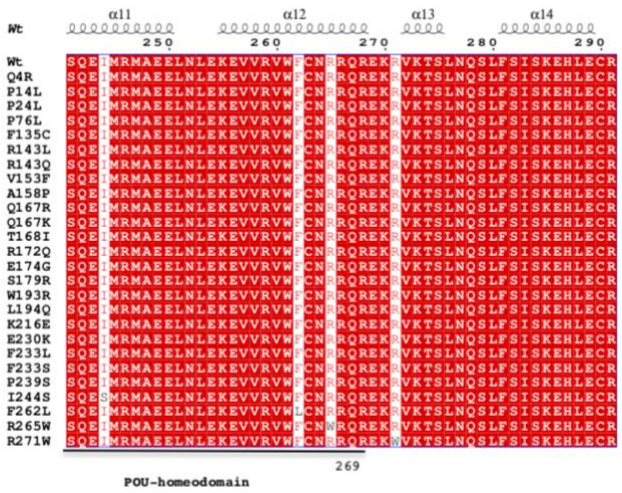


Figure 3. POU1F1 sequence, mutants, and structure.

The missense mutant sequences were aligned with the wildtype sequence structure as topological guide. Every sequence maintains 291 amino acids. Helical structures are portrayed above the sequences and protein functional domains bellow. Note mutation differences in black and white.

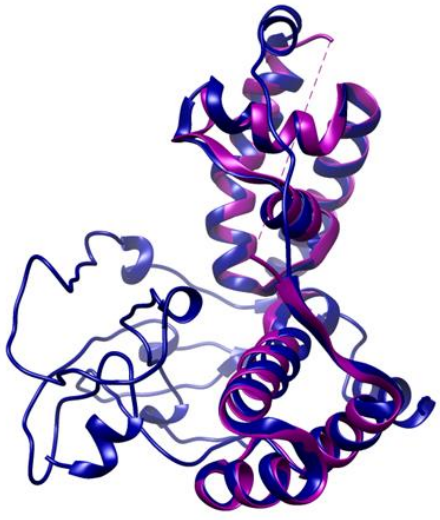


Fig 4A

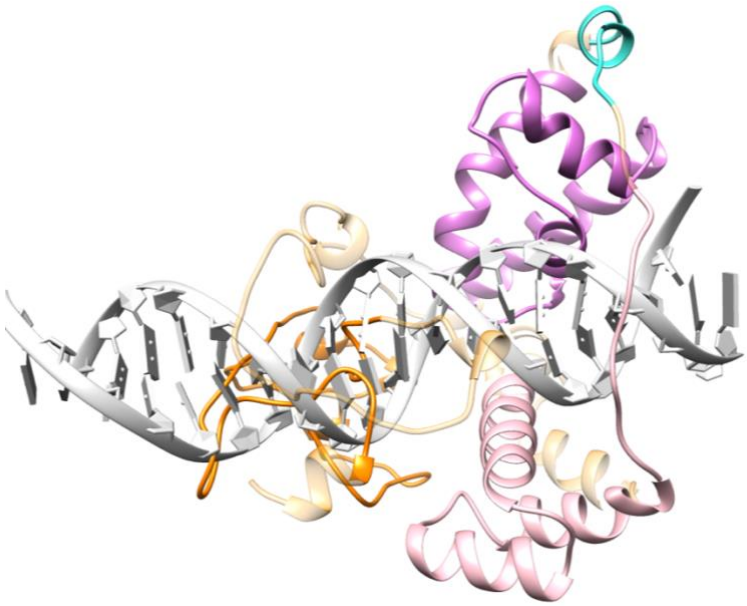


Fig 4B

Figure 4. POU1F1 model alignment to crystallographic structure.

For Figure 4A, the aligned length= 136, Seq_ID=n_identical/n_aligned= 1.000, TM-score= 0.94949 (normalized by the length of the structure modelled with I-TASSER TM-allign). Structure in purple is the one previously modelled by x-ray crystallography (Agarwal & Cho, 2018).The chain in blue is the result of modelling the whole sequence. Figure 4B shows the POU1F1 modelling result binding to the DNA. The colors represent each different domain: Transactivation Domain (TAD) in orange (amino acids 8 to 80), the POU specific (POUS) in purple (AA 127 - 197), linker region (Linker) in turquoise (AA 201 - 209), and POU homeodomain (POUHD) in pink (AA 213 - 269).

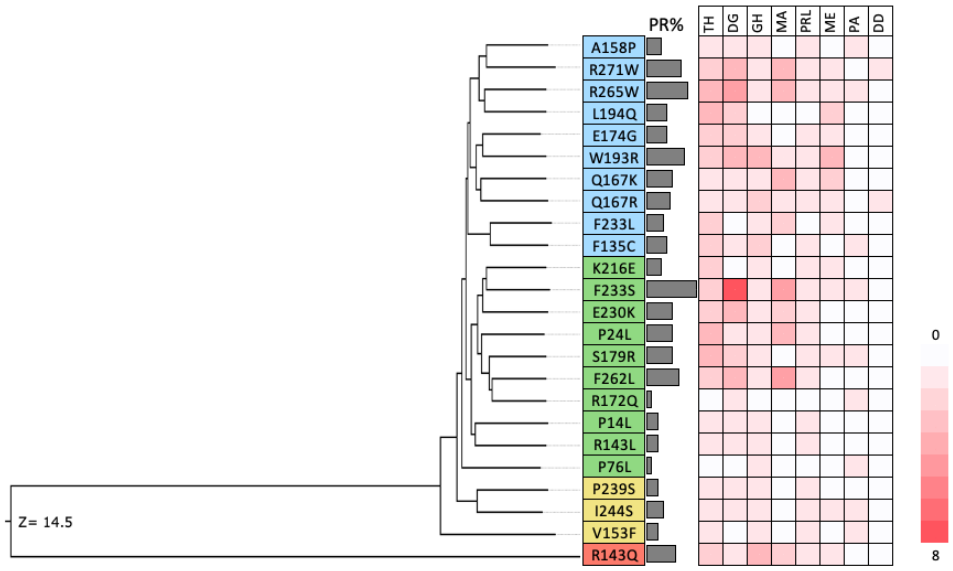


Figure 5. Mutant POU1F1 structural pathological relationships

Four main clusters grouped by their folding performance and sequence resemblance were grouped by colors (orange, yellow, green, and blue). Branches length represent percentage of structural change. The obtained dendrogram was used to infer pathological clusters. In this way pathological matrix was ensemble using an Evo–devo and system pathology logic. Heat map is

arranged from right to left from the most, to the least affected system. Red gradient present pathology diversity range. TH= thyroid hormone deficiency, DG = delayed growth and development, MA= midline abnormalities, PRL = prolactin deficiency, ME= metabolism pathology, PA= pituitary abnormalities, and DD = developmental delay.

6. DISCUSSION

Patients diagnosed with CPHD have variations in their most common clinical manifestations. These variations have been well studied and associated with different mutations in the POU1F1 gene. Based on these studies and available information, this work contributes with the design of a predictive model that associate the varying phenotype of CPHD with the described mutations in groups that would help in the development of better and personalized treatment. The use of bioinformatics allowed the comprehension of how affected could be POU1F1 by its mutations with a more profound understanding than only sequencing and DNA identification. Using bioinformatics allowed us to analyze missense mutations in POU1F1 and study CPHD focusing on the protein's structure and function rather than the DNA sequence.

Naturally, the severity of the patient's phenotype isn't only caused by the mutant protein's functionality and structure. The degree of severity involves the environment and it could also be related to their heritage. Therefore, knowing the patient's ethnicity allows us to identify patterns, so, presented in Figure 1 how the ethnicity relates to the mutations. Mutations in Caucasian patients are mainly shown in exon 4, while exon 6 mutations are mostly present in Asiatic and Arabic mutations (Annex D). Association of this trend with the exons where the mutations occur, showed no stronger evidence to relate the severity to the ethnicity. Nonetheless, there is an informational

bias related to the clinical reports obtained from the SLR regarding the reported cases and how detailed was their case description.

The mutations P24L and R271W are remarkable as examples of mutational hotspots found in POU1F1's DNA sequence. The mutation P24L, as well as P14L and P76L are susceptible to mutations because the DNA sequence mutates from CCT (phenylalanine) to CTT (leucine) (*HGMD® gene result*, n.d.), which may be caused by UV light damage because it has a YY motif (Rogozin & Pavlov, 2003). The mutation P24L has been reported twice, first on a Japanese patient and the second time on a Caucasian patient. The second patient is actually a case of familiar CPHD where the affected relatives also presented lipedema, a symptom that was not present on the Japanese patient (Bano et al., 2010; Ohta et al., 1992). Reports of the same mutation in different parts of the world confirm P24L is a mutational hotspot.

On the other hand, the R271W mutation has a DNA mutation of CGG to TGG. This mutation is similar to the R143L, R143Q, R172Q, and R265W mutations because their arginine codon contains a CG dinucleotide (*HGMD® gene result*, n.d.). The CG dinucleotide which has been long reported to have a high mutation rate, around 35% of DNA mutations occur in a CG dinucleotide (Cooper & Youssoufian, 1988). The R271W mutation has been reported several times, mainly in the United States, but also in Brazil (Birla et al., 2016), Japan (Y. Irie et al., 1995; Ohta K, Nobukuni Y, Mitsubuchi H, Fujimoto S, Matsuo N, Inagaki H, Endo F, 1992; Okamoto et al., 1994), and various European countries (Birla et al., 2016; L. E. Cohen et al., 1995; De Graaff et al., 2010; Holl et al., 1997). In one severe case in particular, the vertical transmission of the mutation was reported in a newborn with CPHD that presented with lack of ossification centers, hyperbilirubinemia, speech delay and gross motor development delay (Taha et al., 2005).

Remarking the importance of genetic testing and prenatal care, especially with dominant mutations in the POU1F1.

Exons 4 and 6, contain several mutational hotspots and also present the most severe phenotype (Figure 2). The severity of their pathology is exemplified by the phenotypes (symptoms) described and their incidence. Exons 4 and 6 are also part of the DNA binding domain of POU1F1, which could explain their most severe phenotype. Interestingly, even though the mutations found in exons 4 and 6 are positioned in mutational hotspots, they are all well conserved amino acids when compared to POU domains present in different transcription factors (Herr et al., 1988b). The fact that these mutations occur in conserved amino acids also explains the patient's severe symptoms.

Mutations that do not occur in the POU domains also impair POU1F1's function, however not as severe but they can be dominant mutations. For example, mutations in the transactivation domain, TAD, are dominant because they compromise the homodimer formation. The TAD is a flexible region that interacts with CBP, but even if this interaction is compromised, there is a greater tolerance for this function, so the phenotype is not as severe. These mutations are also presented in Figure 3 and Annex G. Exons 1 and 2 compose the TAD and its structure is much more flexible than that of the POU$_S$ and POU$_{HD}$. Exon 5 only presents the K216E mutation inside the linker region. The linker which is a flexible structure (Pine-Twaddell et al., 2013) and is required for dimerization, a critical function of POU1F1. This mutation is a recessive mutation and is also causes one of the least severe phenotypes. This leads us to believe that mutations occurring in flexible regions cause less severe patho-phenotype.

To compare the protein structure obtained in our study with the literature, we show in Figure 3 the results obtained from manuscripts showing the DNA binding regions (*DisProt*, n.d.), domains (Aurora & Herr, 1992), and helices (Agarwal & Cho, 2018). The literature has shown evidence for a POU1F1 model with 8 helixes (L. E. Cohen, Hashimoto, et al., 1999) and result of the modelling with I-TASSER, shows a total of 16 helixes. This is likely due to the previously reported structures based on the information provided from the crystallographic x-ray result which only models residues Pit-1$_{\Delta N}$ residues 105–291 (POU$_{HD}$) and Pit-1$_{\Delta N \Delta C}$ residues 124–273 (POU$_S$) (Agarwal & Cho, 2018). The POU$_S$ has been shown to have four helixes, as does our model. The POU$_{HD}$ using x-ray in the crystallography reveals 3 helixes, similar to our model. An additional helix was shown using the crystallographic data and is also seen in our model. also shows a new helix from amino acid 281 to 289.

The three helices we present are present in the TAD, an area of the protein which has not previously been modeled before. Since it is a flexible region, the possible loops that form are described in Table 1(Agarwal & Cho, 2018). Further, it is a serine-threonine rich domain which makes it an intrinsically disordered protein (Sporici et al., 2005). In recent years, computational methods have been developed to predict their structure (Agarwal & Cho, 2018). POU1F1 has a disorder content of 30.3% due to the amino acid content of the TAD and linker regions(He et al., 2009). Using I-TASSER we obtained a structure with three helixes, which are small and may form and may also change since this region is likely to change its structure in order to bind to the different genes. Although I-TASSER is not one of the servers that predicts disordered proteins, it is likely predictive since POU1F1 interaction with p300/CBP is stabilized by the TAD allowing binding to the DNA (*DisProt*, n.d.) and our modelling used POU1F1 binding to the DNA CATT motif (Freedman et al., 2003). Using our new model that can account for the TAD region, may

allow us to be able to model mutant proteins and predict their DNA binding capacity. Further, both I-TASSER and DisEMBL predict that the TAD region is likely to form loops in the described regions (Table 1). These may form under differing conditions hence allowing the protein to interact with various other proteins. This sequence is highly conserved and integrates a regulatory domain and a basal and Ras-responsive region (Agarwal & Cho, 2018), the latter responsible for activating POU1F1 by binding to the CBP/300 complex (Sobrier et al., 2016a).

Structure matching was performed to show the similarity between the protein structure obtained using our modeling and the result from the x-ray crystallographic assay. In Figure 4 the smaller structure is obtained empirically while the structure shown in blue is our result from the server I-TASSER which was performed using the entire 291 amino acid sequence. There were only 136 aligned amino acids which present a TM-score of 0.95 but the alignment of the identical sequence presented scores 1.00, likely due to the last amino acid sequence forming a helix, unlike the crystallographic result. The TM-align modeling result of the mutations when compared with the wildtype structure was not relevant and using Chimera to identify mutations that could present clashes with the structure did not lead to an explanation of how the mutations affect the structure.

Alignment of the mutant proteins with the wildtype template (model 4) showed little difference, with a TM-score above 0.98 in all cases except one (mutation R143Q, TM-score = 0.51). An example of how the modelled *wt* aligns to the crystallographic structure is shown on **Error! Reference source not found.**a. The difference between the mutant structures and the wildtype model is mostly due to the formation of a helical structure in the mutant proteins in amino acids 109 to 118. This helical structure is not present in the wildtype that was used as the template. However, this helical structure was present on a different wildtype model (model 1) obtained from

the initial protein modeling described previously but discarded because of its interaction with the DNA. Alignment of R143Q was low because the construction of the TAD resembled the previously described wildtype model 1 which had a mirror structure of the chosen model to the template.

PyDock analysis was performed to identify physical and chemical differences between the *wt* and the mutant proteins, using the product of the I-TASSER modelling. Understanding that there are two processes about POU1F1 that are well described and are vital to its function, it was decided that the interaction on POU1F1 and DNA should be analyzed. However, both modellings through PyDock were not successful because they didn't present what has been described before as the normal POU1F1 activity. As presented in **Error! Reference source not found.**b, POU1F1 interferes with the DNA, which may cause DNA breakage and it doesn't happen in nature. A similar result was obtained through POU1F1 model interaction with CBP, which naturally binds to the TAD region, but it is not what was seen with PyDock. Even when amino acids were specified for the interactions, the result was not as previously described. Since molecular docking doesn't consider the full flexibility of POU1F1, these unsuccessful results were obtained. Therefore, the binding energies obtained were not further analyzed. A new type of analysis was performed to find the DNA binding energy of POU1F1 using the server SAMPDI (results in Annex F), however there was no evidence or causality with the described phenotype for each mutation.

Hence, evo-devo-med analyses were performed (Binart, 2017). As previously mentioned, dendrogram presents 4 main clusters of neighboring structures. Groups were reconstructed based on differences on folding quantum mechanism and sequence alignment (Figure 5) (Jankowski & Jankowski, 2013)). This allow us to group amino acid changes by their total performance and not only based on an exact change.

Finally, clustering proteins by this quantum relationships presents a more robust confidence, since folding fingerprints will correlate structural isoforms and function alterations in a more correct way. Therefore, the canonical state of POU1F1 can be disrupted by mutations that impact the H+ bonds, electrostatic forces, van der Waals contacts and interactions that results from hydrophobic effect.

All the clusters present as base line a TH and DG disorder, involving distinct pathological phenotypes mainly characterized by lack of TSH, T4 and T3, whereas for DG the symptoms are more varied but generally include musculoskeletal-related pathologies (See Annex E for complete phenotypic characteristic). As consequence of this base TH and DG pathology, different patho-phenotypic patterns arise during development and growth (Figure 5). These patterns allow us to correlate the structural degree of change obtained through the dendrogram and patho-phenotypic characteristics.

The orange outgroup member R143Q show the most different structure compared to all. This mutation causes a flip in the folding patters. In consequence and comparing it to the *wt*, this structure occurs as an optical isomer conformation, having a mirror image when superimposed which gives the 0.54 in similarity with the wildtype (Annex F). As consequence patients presents an optic isomer uncapable of executing an appropriate binding and correct recruitment of the whole transcription machinery (Holm, 2019). This physical impediment is notable according to the chemical-structural dependence of this transcriptional factor (Aykut et al., 2014).

Yellow group of proteins present a cluster of non-groupable proteins since their only relationship to belong is have mutations on non-DNA binding residues neither architectural domains of the protein. Being these secondary structural amino acids or residues that doesn't

perturbed the binding dynamic when performed, presenting a minimum change in structure (Figure 5).

The rest of the two clusters were formed with more specificity depending on the domain affected by the topological changes. Moreover, pathologically both clusters will present representatives in all pathogenicity levels. Blue group involves patients with mutations that only affect functional structures (POUs & POU$_{HD}$). All of the mutants on this group will display changes centralized on the POUs & POU$_{HD}$ domains (Figure 5). In contrast, Green members are affected principally on structural amino acids or secondary functional parts like the TAD. Mutants of this last group are represented as the most variable of structure clusters. Being the consequence changes that occur on non-catalytical or support structures. In this way, structures with an extreme importance that result affected by more drastically changes will present more pathogenic phenotypes (Figure 5).

F233S presents the most pathogenic phenotype of all, their affection is highly spread among systems, only DD presents no deficiency, however this is likely due to the early death of the patient (Baş et al., 2015b). The localization of this mutation is in the tenth alfa helix, correspond to the POU-homeodomain. This par in specific don't directly bind to the DNA, but the change of one highly hydrophobic amino acid, the phenylalanine (F) by a serine (S) with polar uncharged side chains makes a big difference on helix.

7. CONCLUSION

In conclusion, a more comprehensive way to interpret pathological phenotypes could be based on protein structural changes. This could allow us, and future clinicians on a better comprehension of the patient's evolution trough time. Since, most of these pathologies affect the patients thought their lifetime. Our analysis helps us in this way to link this structural dependence on the different levels of pathological phenotypes that patients could exhibit. Moreover, structural differences are not the only energetic characteristic for base on a final distribution. Therefore, we propose that a molecular dynamics analysis of the DNA binding could further explain the clinical phenotypes. Dimerization, phosphorylation and CBP binding energies could also be analyzed in further researches.

The published data about each patient and their mutations was generally very detailed which allowed the construction of the phylogenetic tree of the mutations. The bioinformatic tools used didn't account for the flexibility of the protein, which could be why we weren't able to relate the protein structure to the pathological phenotypes. Therefore, we suggest that the bioinformatic tools used in future studies involve molecular dynamics and take on account the following aspects: DNA binding, CBP binding to TAD, and phosphorylation by PKA.

8. BIBLIOGRAPHY

Aarskog, D., Eiken, H. G., Bjerknes, R., & Myking, O. L. (1997). Pituitary dwarfism in the R271W Pit-1 gene mutation. *European Journal of Pediatrics*. https://doi.org/10.1007/s004310050722

Agarwal, S., & Cho, T. Y. (2018). Biochemical and structural characterization of a novel cooperative binding mode by Pit-1 with CATT repeats in the macrophage migration inhibitory factor promoter. *Nucleic Acids Research*, *46*(2), 929–941. https://doi.org/10.1093/nar/gkx1183

Augustijn, K. D., Bakker, B., Holthuizen, P., Van Der Vliet, P. C., & Jansen, M. (2001). Combined pituitary hormone deficiency caused by compound heterozygosity for two novel mutations in the POU domain of the PIT1/POU1F1 gene. *Journal of Clinical Endocrinology and Metabolism*, *86*(4), 1545–1550. https://doi.org/10.1210/jc.86.4.1545

Aurora, R., & Herr, W. (1992). Segments of the POU domain influence one another's DNA-binding specificity. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.12.2.455

Aykut, A., Ozen, S., Simsek, D. G., Onay, H., Cogulu, O., Darcan, S., & Ozkinay, F. (2014). Multiple cutaneous hemangiomas in a patient with combined pituitary hormone deficiency. *Journal of Pediatric Endocrinology and Metabolism*, *27*(1–2), 171–174. https://doi.org/10.1515/jpem-2013-0208

Bano, G., Mansour, S., Brice, G., Ostergaard, P., Mortimer, P. S., Jeffery, S., & Nussey, S. (2010). Pit-1 mutation and lipoedema in a family. *Experimental and Clinical Endocrinology and Diabetes*, *118*(6), 377–380. https://doi.org/10.1055/s-0029-1224154

Baş, F., Abalı, Z. Y., Toksoy, G., Poyrazoğlu, Ş., Bundak, R., Güleç, Ç., Uyguner, Z. O., &

Darendeliler, F. (2018). Precocious or early puberty in patients with combined pituitary hormone deficiency due to POU1F1 gene mutation: case report and review of possible mechanisms. *Hormones (Athens, Greece)*, *17*(4), 581–588. https://doi.org/10.1007/s42000-018-0079-4

Baş, F., Uyguner, Z. O., Darendeliler, F., Aycan, Z., Çetinkaya, E., Berberoğlu, M., Şiklar, Z., Öcal, G., Darcan, Ş., Gökşen, D., Topaloğlu, A. K., Yüksel, B., Özbek, M. N., Ercan, O., Evliyaoğlu, O., Çetinkaya, S., Şen, Y., Atabek, E., Toksoy, G., … Bundak, R. (2015a). Molecular analysis of PROP1, POU1F1, LHX3, and HESX1 in Turkish patients with combined pituitary hormone deficiency: a multicenter study. *Endocrine*. https://doi.org/10.1007/s12020-014-0498-1

Baş, F., Uyguner, Z. O., Darendeliler, F., Aycan, Z., Çetinkaya, E., Berberoğlu, M., Şiklar, Z., Öcal, G., Darcan, Ş., Gökşen, D., Topaloğlu, A. K., Yüksel, B., Özbek, M. N., Ercan, O., Evliyaoğlu, O., Çetinkaya, S., Şen, Y., Atabek, E., Toksoy, G., … Bundak, R. (2015b). Molecular analysis of PROP1, POU1F1, LHX3, and HESX1 in Turkish patients with combined pituitary hormone deficiency: a multicenter study. *Endocrine*, *49*(2), 479–491. https://doi.org/10.1007/s12020-014-0498-1

Bastos, E., Santos, I., Parmentier, I., Castrillo, J. L., Cravador, A., Guedes-Pinto, H., & Renaville, R. (2006). Ovis aries POU1F1 gene: Cloning, characterization and polymorphism analysis. *Genetica*. https://doi.org/10.1007/s10709-005-0034-6

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-09654-8

Binart, N. (2017). Chapter 5 – Prolactin. In *The Pituitary*. https://doi.org/10.1016/B978-0-12-

804169-7.00005-2

Bircan, I., Semiz, S., Cogan, J. D., Hedges, L. K., & Semiz, S. (2001). Complete Deficiency of Growth Hormone, Thyrotropin and Prolactin in Two Siblings: A New Pit-1 Mutation. *International Journal on Disability and Human Development*, *2*(1), 61–68. https://doi.org/10.1515/IJDHD.2001.2.1.61

Birla, S., Khadgawat, R., Jyotsna, V. P., Jain, V., Garg, M. K., Bhalla, A. S., & Sharma, A. (2016). Identification of Novel PROP1 and POU1F1 Mutations in Patients with Combined Pituitary Hormone Deficiency. *Hormone and Metabolic Research*. https://doi.org/10.1055/s-0042-117112

Blum, W. F., Klammt, J., Amselem, S., Pfäffle, H. M., Legendre, M., Sobrier, M. L., Luton, M. P., Child, C. J., Jones, C., Zimmermann, A. G., Quigley, C. A., Cutler, G. B., Deal, C. L., Lebl, J., Rosenfeld, R. G., Parks, J. S., & Pfäffle, R. W. (2018a). Screening a large pediatric cohort with GH deficiency for mutations in genes regulating pituitary development and GH secretion: Frequencies, phenotypes and growth outcomes. *EBioMedicine*. https://doi.org/10.1016/j.ebiom.2018.09.026

Blum, W. F., Klammt, J., Amselem, S., Pfäffle, H. M., Legendre, M., Sobrier, M. L., Luton, M. P., Child, C. J., Jones, C., Zimmermann, A. G., Quigley, C. A., Cutler, G. B., Deal, C. L., Lebl, J., Rosenfeld, R. G., Parks, J. S., & Pfäffle, R. W. (2018b). Screening a large pediatric cohort with GH deficiency for mutations in genes regulating pituitary development and GH secretion: Frequencies, phenotypes and growth outcomes. *EBioMedicine*, *36*, 390–400. https://doi.org/10.1016/j.ebiom.2018.09.026

Brown, M. R., Parks, J. S., Adess, M. E., Rich, B. H., Rosenthal, I. M., Voss, T. C., VanderHeyden, T. C., & Hurley, D. L. (1998). Central hypothyroidism reveals compound heterozygous

mutations in the Pit-1 gene. *Hormone Research*, *49*(2), 98–102. https://doi.org/10.1159/000023134

Cao, Y., Siegel, D., Donow, C., Knöchel, S., Yuan, L., & Knöchel, W. (2007). POU-V factors antagonize maternal VegT activity and β-Catenin signaling in Xenopus embryos. *EMBO Journal*. https://doi.org/10.1038/sj.emboj.7601736

Carlomagno, Y., Salerno, M., Vivenza, D., Capalbo, D., Godi, M., Mellone, S., Tiradani, L., Corneli, G., Momigliano-Richiardi, P., Bona, G., & Giordano, M. (2009). A novel recessive splicing mutation in the POU1F1 gene causing combined pituitary hormone deficiency. *Journal of Endocrinological Investigation*, *32*(8), 653–658. https://doi.org/10.1007/bf03345736

Cheng, T. M. K., Blundell, T. L., & Fernandez-Recio, J. (2007). PyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins: Structure, Function and Genetics*. https://doi.org/10.1002/prot.21419

Cohen, L. E., Hashimoto, Y., Zanger, K., Wondisford, F., & Radovick, S. (1999). CREB-independent regulation by CBP is a novel mechanism of human growth hormone gene expression. *Journal of Clinical Investigation*. https://doi.org/10.1172/jci7308

Cohen, L. E., Wondisford, F. E., Salvatoni, A., Maghnie, M., Brucker-Davis, F., Weintraub, B. D., & Radovick, S. (1995). A “hot spot” in the Pit-1 gene responsible for combined pituitary hormone deficiency: Clinical and molecular correlates. *Journal of Clinical Endocrinology and Metabolism*, *80*(2), 679–684. https://doi.org/10.1210/jcem.80.2.7852536

Cohen, L. E., Zanger, K., Brue, T., Wondisford, F. E., & Radovick, S. (1999). Defective retinoic acid regulation of the Pit-1 gene enhancer: A novel mechanism of combined pituitary hormone deficiency. *Molecular Endocrinology*, *13*(3), 476–484.

https://doi.org/10.1210/mend.13.3.0251

Cohen, R. N., Brue, T., Naik, K., Houlihan, C. A., Wondisford, F. E., & Radovick, S. (2006). The role of CBP/p300 interactions and Pit-1 dimerization in the pathophysiological mechanism of combined pituitary hormone deficiency. *Journal of Clinical Endocrinology and Metabolism*, *91*(1), 239–247. https://doi.org/10.1210/jc.2005-1211

Combined pituitary hormone deficiency - Genetics Home Reference - NIH. (n.d.). Retrieved March 10, 2020, from https://ghr.nlm.nih.gov/condition/combined-pituitary-hormone-deficiency#sourcesforpage

Cooper, D. N., & Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. *Human Genetics*, *78*(2), 151–155. https://doi.org/10.1007/BF00278187

De Graaff, L. C. G., Argente, J., Veenma, D. C. M., Drent, M. L., Uitterlinden, A. G., & Hokken-Koelega, A. C. S. (2010). PROP1, HESX1, POU1F1, LHX3 and LHX4 mutation and deletion screening and GH1 P89L and IVS3+1/+2 mutation screening in a dutch nationwide cohort of patients with combined pituitary hormone deficiency. *Hormone Research in Paediatrics*, *73*(5), 363–371. https://doi.org/10.1159/000308169

De Rienzo, F., Mellone, S., Bellone, S., Babu, D., Fusco, I., Prodam, F., Petri, A., Muniswamy, R., De Luca, F., Salerno, M., Momigliano-Richardi, P., Bona, G., & Giordano, M. (2015). Frequency of genetic defects in combined pituitary hormone deficiency: A systematic review and analysis of a multicentre Italian cohort. *Clinical Endocrinology*. https://doi.org/10.1111/cen.12849

DisProt. (n.d.). Retrieved March 10, 2020, from https://www.disprot.org/DP00231?release=current&show_ambiguous=true&show_obsolete=false

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C. H., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, *19*(1), 26–59. https://doi.org/10.1016/S1093-3263(00)00138-8

Elizabeth, M., Hokken-Koelega, A. C. S., Schuilwerve, J., Peeters, R. P., Visser, T. J., & de Graaff, L. C. G. (2018). Genetic screening of regulatory regions of pituitary transcription factors in patients with idiopathic pituitary hormone deficiencies. *Pituitary*. https://doi.org/10.1007/s11102-017-0850-6

Fang, Q., George, A. S., Brinkmeier, M. L., Mortensen, A. H., Gergics, P., Cheung, L. Y. M., Daly, A. Z., Ajmal, A., Millán, M. I. P., Bilge Ozel, A., Kitzman, J. O., Mills, R. E., Li, J. Z., & Camper, S. A. (2016). Genetics of combined pituitary hormone deficiency: Roadmap into the genome era. In *Endocrine Reviews*. https://doi.org/10.1210/er.2016-1101

Fofanova, O. V., Takamura, N., Kinoshita, E. I., Yoshimoto, M., Tsuji, Y., Peterkova, V. A., Evgrafov, O. V., Dedov, I. I., Goncharov, N. P., & Yamashita, S. (1998). Rarity of PIT1 involvement in children from Russia with combined pituitary hormone deficiency. *American Journal of Medical Genetics*, *77*(5), 360–365. https://doi.org/10.1002/(SICI)1096-8628(19980605)77:5<360::AID-AJMG4>3.0.CO;2-R

Freedman, S. J., Sun, Z. Y. J., Kung, A. L., France, D. S., Wagner, G., & Eck, M. J. (2003). Structural basis for negative regulation of hypoxia-inducible factor-1α by CITED2. *Nature Structural Biology*. https://doi.org/10.1038/nsb936

Garner, Romero, Dunker, Brown, & Obradovic. (1999). Predicting Binding Regions within Disordered Proteins. *Genome Informatics. Workshop on Genome Informatics*.

https://doi.org/10.11234/gi1990.10.41

Gat-Yablonski, G., Klar, A., Hirsch, D., Eliakim, A., Lazar, L., Hurvitz, H., & Phillip, M. (2005). Three novel mutations in POU1F1 in Israeli patients with combined pituitary hormone deficiency. *Journal of Pediatric Endocrinology and Metabolism*, *18*(4), 385–393. https://doi.org/10.1515/JPEM.2005.18.4.385

Gat-Yablonski, Galia, David, R. F. Ben, Bar, M., Potievsky, O., Phillip, M., & Lazar, L. (2011). Homozygous microdeletion of the POU1F1, CHMP2B, and VGLL3 genes in chromosome 3-A novel syndrome. *American Journal of Medical Genetics, Part A*. https://doi.org/10.1002/ajmg.a.34136

Gavrilova, A. E., Nagaeva, E. V., Rebrova, O. Y., Shiryaeva, T. Y., Tiulpakov, A. N., & Peterkova, V. A. (2017). A novel heterozygous mutation in POU1F1 is associated with combined pituitary hormone deficiency. *Problemy Endokrinologii*. https://doi.org/10.14341/probl2017633169-173

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, *30*(SUPPL. 1). https://doi.org/10.1002/elps.200900140

Hashimoto, Y., Cisternino, M., & Cohen, L. E. (2003). A novel nonsense mutation in the Pit-1 gene: Evidence for a gene dosage effect. *Journal of Clinical Endocrinology and Metabolism*, *88*(3), 1241–1247. https://doi.org/10.1210/jc.2002-021510

Hatos, A., Hajdu-Soltész, B., Monzon, A. M., Palopoli, N., Álvarez, L., Aykac-Fas, B., Bassot, C., Benítez, G. I., Bevilacqua, M., Chasapi, A., Chemes, L., Davey, N. E., Davidović, R., Dunker, A. K., Elofsson, A., Gobeill, J., Foutel, N. S. G., Sudha, G., Guharoy, M., … Piovesan, D. (2020). DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Research*.

https://doi.org/10.1093/nar/gkz975

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., & Dunker, A. K. (2009). Predicting intrinsic disorder in proteins: An overview. In *Cell Research*. https://doi.org/10.1038/cr.2009.87

Hendriks-Stegeman, B. I., Augustijn, K. D., Bakker, B., Holthuizen, P., van der Vliet, P. C., & Jansen, M. (2001). Combined Pituitary Hormone Deficiency Caused by Compound Heterozygosity for Two Novel Mutations in the POU Domain of the PIT1/POU1F1 Gene 1 . *The Journal of Clinical Endocrinology & Metabolism*, *86*(4), 1545–1550. https://doi.org/10.1210/jcem.86.4.7371

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., & Ruvkun, G. (1988a). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. *Genes & Development*. https://doi.org/10.1101/gad.2.12a.1513

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., & Ruvkun, G. (1988b). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. *Genes & Development*, *2*(12 A), 1513–1516. https://doi.org/10.1101/gad.2.12a.1513

HGMD® gene result. (n.d.). Retrieved March 10, 2020, from http://www.hgmd.cf.ac.uk/ac/gene.php?gene=POU1F1

Holl, R. W., Pfäffle, R., Kim, C., Sorgo, W., Teller, W. M., & Heimann, G. (1997). Combined pituitary deficiencies of growth hormone, thyroid stimulating hormone and prolactin due to Pit-1 gene mutation: A case-report. *European Journal of Pediatrics*. https://doi.org/10.1007/s004310050723

Holm, L. (2019). Benchmarking fold detection by DaliLite v.5. *Bioinformatics (Oxford, England)*. https://doi.org/10.1093/bioinformatics/btz536

Hroudova, M., Vojta, P., Strnad, H., Krejcik, Z., Ridl, J., Paces, J., Vlcek, C., & Paces, V. (2012). Diversity, phylogeny and expression patterns of pou and six homeodomain transcription factors in hydrozoan jellyfish craspedacusta sowerbyi. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0036420

Inoue, H., Mukai, T., Sakamoto, Y., Kimura, C., Kangawa, N., Itakura, M., Ogata, T., Ito, Y., & Fujieda, K. (2012). Identification of a novel mutation in the exon 2 splice donor site of the POU1F1/PIT-1 gene in Japanese identical twins with mild combined pituitary hormone deficiency. *Clinical Endocrinology*, *76*(1), 78–87. https://doi.org/10.1111/j.1365-2265.2011.04165.x

Irie, Y., Tatsumi, K. I., Kusuda, S., Kawawaki, H., Boyages, S. C., Nose, O., Ichiba, Y., Katsumata, N., & Amino, N. (1995). Screening for PIT1 abnormality by PCR direct sequencing method. *Thyroid*. https://doi.org/10.1089/thy.1995.5.207

Irie, Yoshifumi, Tatsumi, K. ita, Amino, N., Ogawa, M., Kamijo, T., Preeyasombat, C., & Suprasongsin, C. (1995). A Novel E250X Mutation of the Pit1 Genein a Patient with Combined Pituitary Hormone Deficiency. *Endocrine Journal*. https://doi.org/10.1507/endocrj.42.351

Jankowski, R., & Jankowski, R. (2013). Evolutionary and Developmental (Evo-Devo) Medicine. In *The Evo-Devo Origin of the Nose, Anterior Skull Base and Midface* (pp. 199–202). Springer Paris. https://doi.org/10.1007/978-2-8178-0422-4_20

Jasper J. Koehorst, Jesse C. J. van Dam, Edoardo Saccenti, Vitor A. P. Martins dos Santos, M. S.-D. and P. J. S. (2017). Imported from

https://www.researchgate.net/publication/321428915_SAPP_functional_genome_annotation_and_analysis_through_a_semantic_framework_using_FAIR_principles. *Oxford*, *34*(8), 1401–1403. https://doi.org/10.1093/BIOINFORMATICS

Johnson, W. A., & Hirsh, J. (1990). Binding of a Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. *Nature*. https://doi.org/10.1038/343467a0

Karp, G. (2009). Cell and Molecular Biology: Concepts and Experiments. *Wiley; 6th Edition*.

Kinoshita, E., Parks, J. S., & Brown, M. R. (1994). *Splicing with Variation of PIT-1 TSH mRNA Deficiency in*. *3*(Suppl 4), 77–82.

Lee, N. C., Tsai, W. Y., Peng, S. F., Tung, Y. C., Chien, Y. H., & Hwu, W. L. (2011). Congenital hypopituitarism due to POU1F1 gene mutation. *Journal of the Formosan Medical Association*, *110*(1), 58–61. https://doi.org/10.1016/S0929-6646(11)60009-0

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003a). Protein disorder prediction: Implications for structural proteomics. *Structure*. https://doi.org/10.1016/j.str.2003.10.002

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003b). Protein disorder prediction: Implications for structural proteomics. *Structure*, *11*(11), 1453–1459. https://doi.org/10.1016/j.str.2003.10.002

Malvagia, S., Poggi, G. M., Pasquini, E., Donati, M. A., Pela, I., Morrone, A., & Zammarchi, E. (2003). The De Novo Q167K Mutation in the POU1F1 Gene Leads to Combined Pituitary Hormone Deficiency in an Italian Patient. *Pediatric Research*, *54*(5), 635–640. https://doi.org/10.1203/01.PDR.0000084113.41375.1E

McLennan, K., Jeske, Y., Cotterill, A., Cowley, D., Penfold, J., Jones, T., Howard, N., Thomsett,

M., & Choong, C. (2003). Combined pituitary hormone deficiency in Australian children: Clinical and genetic correlates. *Clinical Endocrinology*, *58*(6), 785–794. https://doi.org/10.1046/j.1365-2265.2003.01781.x

Miyata, I., Vallette-Kasic, S., Saveanu, A., Takeuchi, M., Yoshikawa, H., Tajima, A., Tojo, K., Reynaud, R., Gueydan, M., Enjalbert, A., Tajima, N., Eto, Y., & Brue, T. (2006). Identification and functional analysis of the novel S179R POU1F1 mutation associated with combined pituitary hormone deficiency. *Journal of Clinical Endocrinology and Metabolism*, *91*(12), 4981–4987. https://doi.org/10.1210/jc.2005-2289

Ohta K, Nobukuni Y, Mitsubuchi H, Fujimoto S, Matsuo N, Inagaki H, Endo F, M. I. (1992). *MUTATIONS IN THE PIT-l GENE IN CHILDREN wITH COMBINED PITUITARY HORMONE DEFICIENCY*. *157*(2), 1063–1068.

Ohta, K., Nobukuni, Y., Mitsubuchi, H., Fujimoto, S., Matsuo, N., Inagaki, H., Endo, F., & Matsuda, I. (1992). Mutations in the PIT-1 gene in children with combined pituitary hormone deficiency. *Biochemical and Biophysical Research Communications*, *189*(2), 851–855. https://doi.org/10.1016/0006-291X(92)92281-2

Okamoto, N., Wada, Y., Ida, S., Koga, R., Ozono, K., Chiyo, H. aki, Hayashi, A., & Tatsumi, K. ita. (1994). Monoallelic expression of normal mRNA in the PIT1 mutation heterozygotes with normal phenotype and biallelic expression in the abnormal phenotype. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/3.9.1565

Paper, O. (2002). A New C-Terminal Located Mutation (V272ter) in the PIT-1 Gene Manifesting with. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, *189*(2), 81–86.

Pellegrini-Bouiller, I., Bélicar, P., Barlier, A., Gunz, G., Charvet, J. P., Jaquet, P., Brue, T.,

Vialettes, B., & Enjalbert, A. (2008). PIT-l MUTATION affected children had been found before 3 yr of age to have severe growth retardation associated with undetectable plasma GH and PRL levels and later developed central hy- pothyroidism . We report in the present paper a summary of the clini. *Methods*, *November*.

Pernasetti, F., Milner, R. D. G., Al Ashwal, A. A. Z., De Zegher, F., Chavez, V. M., Muller, M., & Martial, J. A. (1998). Pro239Ser: A novel recessive mutation of the Pit-1 gene in seven Middle Eastern children with growth hormone, prolactin, and thyrotropin deficiency. *Journal of Clinical Endocrinology and Metabolism*, *83*(6), 2079–2083. https://doi.org/10.1210/jc.83.6.2079

Pernasetti, F., Wera, S., Belayew, A., & Martial, J. A. (1993). Cloning of a human GHF-1/PIT-1 cDNA variant. *Nucleic Acids Research*, *21*(15), 3584. https://doi.org/10.1093/nar/21.15.3584

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pfäffle, R. W., DiMattia, G. E., Parks, J. S., Brown, M. R., Wit, J. M., Jansen, M., Van Der Nat, H., Van Den Brande, J. L., Rosenfeld, M. G., & Ingraham, H. A. (1992). Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. *Science*. https://doi.org/10.1126/science.257.5073.1118

Pine-Twaddell, E., Romero, C. J., & Radovick, S. (2013). Vertical transmission of hypopituitarism: Critical importance of appropriate interpretation of thyroid function tests and levothyroxine therapy during pregnancy. *Thyroid*, *23*(7), 892–897. https://doi.org/10.1089/thy.2012.0332

Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C. J., Aspromonte, M. C., Davey, N. E., Davidović, R., Dosztányi, Z., Elofsson, A., Gasparini, A., Hatos, A., Kajava, A. V., Kalmar, L., Leonardi, E., Lazar, T., Macedo-Ribeiro, S., Macossay-Castillo, M., … Tosatto, S. C. E. (2017). DisProt 7.0: A major update of the database of disordered proteins. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw1056

POU1F1 - Pituitary-specific positive transcription factor 1 - Homo sapiens (Human) - POU1F1 gene & protein. (n.d.). Retrieved March 10, 2020, from https://www.uniprot.org/uniprot/P28069

Radovick, S., Nations, M., Du, Y., Berg, L. V. A., Weintraub, B. D., & Wondisford, F. E. (1992). A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science*. https://doi.org/10.1126/science.257.5073.1115

Rainbow, L. A., Rees, S. A., Shaikh, M. G., Shaw, N. J., Cole, T., Barrett, T. G., & Kirk, J. M. W. (2005). Mutation analysis of POUF-1, PROP-1 and HESX-1 show low frequency of mutations in children with sporadic forms of combined pituitary hormone deficiency and septo-optic dysplasia. *Clinical Endocrinology*, *62*(2), 163–168. https://doi.org/10.1111/j.1365-2265.2004.02189.x

Rambaut, A., & FigTree, V. (2018). *1.4. 4*.

Reynaud, R., Gueydan, M., Saveanu, A., Vallette-Kasic, S., Enjalbert, A., Brue, T., & Barlier, A. (2006). Genetic screening of combined pituitary hormone deficiency: Experience in 195 patients. *Journal of Clinical Endocrinology and Metabolism*. https://doi.org/10.1210/jc.2005-2173

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*. https://doi.org/10.1093/nar/gku316

Rogozin, I. B., & Pavlov, Y. I. (2003). Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutation Research - Reviews in Mutation Research*, *544*(1), 65–85. https://doi.org/10.1016/S1383-5742(03)00032-2

Sadeghi-Nejad, A., & Senior, B. (1974). A familial syndrome of isolated "aplasia" of the anterior pituitary. *The Journal of Pediatrics*, *84*(1), 79–84. https://doi.org/10.1016/s0022-3476(74)80557-3

Snabboon, T., Plengpanich, W., Buranasupkajorn, P., Khwanjaipanich, R., Vasinanukorn, P., Suwanwalaikorn, S., Khovidhunkit, W., & Shotelersuk, V. (2007). A novel germline mutation, IVS4+1G>A, of the POU1F1 gene underlying combined pituitary hormone deficiency. *Hormone Research*, *69*(1), 60–64. https://doi.org/10.1159/000111797

Sobrier, M. L., Tsai, Y. C., Pérez, C., Leheup, B., Bouceba, T., Duquesnoy, P., Copin, B., Sizova, D., Penzo, A., Stanger, B. Z., Cooke, N. E., Liebhaber, S. A., & Amselem, S. (2016a). Functional characterization of a human POU1F1 mutation associated with isolated growth hormone deficiency: A novel etiology for IGHD. *Human Molecular Genetics*, *25*(3), 472–483. https://doi.org/10.1093/hmg/ddv486

Sobrier, M. L., Tsai, Y. C., Pérez, C., Leheup, B., Bouceba, T., Duquesnoy, P., Copin, B., Sizova, D., Penzo, A., Stanger, B. Z., Cooke, N. E., Liebhaber, S. A., & Amselem, S. (2016b). Functional characterization of a human POU1F1 mutation associated with isolated growth hormone deficiency: A novel etiology for IGHD. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/ddv486

Sporici, R. A., Hodskins, J. S., Locasto, D. M., Meszaros, L. B., Ferry, A. L., Weidner, A. M., Rinehart, C. A., Bailey, J. C., Mains, I. M., & Diamond, S. E. (2005). Repression of the prolactin promoter: A functional consequence of the heterodimerization between Pit-1 and

Pit-1 β. *Journal of Molecular Endocrinology*, *35*(2), 317–331. https://doi.org/10.1677/jme.1.01678

Stenson, P. D., Mort, M., Ball, E. V., Shaw, K., Phillips, A. D., & Cooper, D. N. (2014). The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. In *Human Genetics*. https://doi.org/10.1007/s00439-013-1358-4

Sun, Y., Zhang, F., Gao, J., Gao, X., Guo, T., Zhang, K., Shi, Y., Zheng, Z., Tang, W., Zheng, Y., Li, S., Li, X., Feng, G., Shen, X., & He, L. (2006). Positive association between POU1F1 and mental retardation in young females in the Chinese Han population. *Human Molecular Genetics*, *15*(7), 1237–1243. https://doi.org/10.1093/hmg/ddl039

Taha, D., Mullis, P. E., Ibáñez, L., & De Zegher, F. (2005). Absent or delayed adrenarche in Pit-1/POU1F1 deficiency. *Hormone Research*, *64*(4), 175–179. https://doi.org/10.1159/000088793

Tenenbaum-Rakover, Y., Sobrier, M. L., & Amselem, S. (2011). A novel POU1F1 mutation (p.Thr168IlefsX7) associated with an early and severe form of combined pituitary hormone deficiency: Functional analysis and follow-up from infancy to adulthood. *Clinical Endocrinology*. https://doi.org/10.1111/j.1365-2265.2011.04028.x

Turton, J. P. G., Reynaud, R., Mehta, A., Torpiano, J., Saveanu, A., Woods, K. S., Tiulpakov, A., Zdravkovic, V., Hamilton, J., Attard-Montalto, S., Parascandalo, R., Vella, C., Clayton, P. E., Shalet, S., Barton, J., Brue, T., & Dattani, M. T. (2005a). Novel mutations within the POU1F1 gene associated with variable combined pituitary hormone deficiency. *Journal of Clinical Endocrinology and Metabolism*, *90*(8), 4762–4770. https://doi.org/10.1210/jc.2005-0570

Turton, J. P. G., Reynaud, R., Mehta, A., Torpiano, J., Saveanu, A., Woods, K. S., Tiulpakov, A.,

Zdravkovic, V., Hamilton, J., Attard-Montalto, S., Parascandalo, R., Vella, C., Clayton, P. E., Shalet, S., Barton, J., Brue, T., & Dattani, M. T. (2005b). Novel mutations within the POU1F1 gene associated with variable combined pituitary hormone deficiency. *Journal of Clinical Endocrinology and Metabolism*, *90*(8), 4762–4770. https://doi.org/10.1210/jc.2005-0570

Turton, J. P., Strom, M., Langham, S., Dattani, M. T., & Le Tissier, P. (2012). Two novel mutations in the POU1F1 gene generate null alleles through different mechanisms leading to combined pituitary hormone deficiency. *Clinical Endocrinology*, *76*(3), 387–393. https://doi.org/10.1111/j.1365-2265.2011.04236.x

UniProtKB/SwissProt variant VAR_049361. (n.d.). Retrieved March 10, 2020, from https://web.expasy.org/variant_pages/VAR_049361.html

Wärnmark, A., Treuter, E., Wright, A. P. H., & Gustafsson, J. Å. (2003). Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation. In *Molecular Endocrinology*. https://doi.org/10.1210/me.2002-0384

Yang, J., Wang, Y., & Zhang, Y. (2016). ResQ: An Approach to Unified Estimation of B-Factor and Residue-Specific Error in Protein Structure Prediction. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2015.09.024

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv342

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., Goddard, T. D., Meng, E. C., Sali, A., & Ferrin, T. E. (2012). UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2011.09.006

Zhang, C., Freddolino, P. L., & Zhang, Y. (2017). COFACTOR: Improved protein function

prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx366

9. ANNEX

Annex A: Search terms

Items	*Search terms*
Protein	((TITLE-ABS-KEY (*POU1F1*)) OR (TITLE-ABS-KEY (GHF-1)) OR (TITLE-ABS-KEY (*Pit-1*)))
Population	((TITLE-ABS-KEY (human*)) OR (TITLE-ABS-KEY (homo sapiens)) OR (TITLE-ABS-KEY (person*)) OR (TITLE-ABS-KEY (people)))
Characteristics	((TITLE-ABS-KEY (structure)) OR (TITLE-ABS-KEY (aminoacid*)) OR (TITLE-ABS-KEY (peptid*)) OR (TITLE-ABS-KEY (Protein AND structure)) OR (TITLE-ABS-KEY (*sequence*)) OR (TITLE-ABS-KEY (proteom*)) OR (TITLE-ABS-KEY (transcript*)) OR (TITLE-ABS-KEY (exome*)) OR (TITLE-ABS-KEY (exon*)) OR (TITLE-ABS-KEY (evolution)) OR(TITLE-ABS-KEY (mut*)) OR (TITLE-ABS-KEY (homolog*)) OR (TITLE-ABS-KEY (paralog*)) OR (TITLE-ABS-KEY (gen*)) OR (TITLE-ABS-KEY (molecular AND modeling)) OR (TITLE-ABS-KEY (splicing)) OR (TITLE-ABS-KEY (functional AND prediction)) OR (TITLE-ABS-KEY (DNA AND binding)))
TOTAL	((TITLE-ABS-KEY (*POU1F1*)) OR (TITLE-ABS-KEY (GHF-1)) OR (TITLE-ABS-KEY (*Pit-1*)))AND ((TITLE-ABS-KEY (human*)) OR (TITLE-ABS-KEY (homo sapiens)) OR (TITLE-ABS-KEY (person*)) OR (TITLE-ABS-KEY (people))) AND ((TITLE-ABS-KEY (structure)) OR (TITLE-ABS-KEY (aminoacid*)) OR (TITLE-ABS-KEY (peptid*)) OR (TITLE-ABS-KEY (Protein AND structure)) OR (TITLE-ABS-KEY (*sequence*)) OR (TITLE-ABS-KEY (proteom*)) OR (TITLE-ABS-KEY (transcript*)) OR (TITLE-ABS-KEY (exome*)) OR (TITLE-ABS-KEY (exon*)) OR (TITLE-ABS-KEY (evolution)) OR(TITLE-ABS-KEY (mut*)) OR (TITLE-ABS-KEY (homolog*)) OR (TITLE-ABS-KEY (paralog*)) OR (TITLE-ABS-KEY (gen*)) OR (TITLE-ABS-KEY (molecular AND modeling)) OR (TITLE-ABS-KEY (splicing)) OR (TITLE-ABS-KEY (functional AND prediction)) OR (TITLE-ABS-KEY (DNA AND binding)))

Annex B: Article classification

Category	*Exclusion criteria*
0. Null entries, duplicates, not in the language of interest, abstract is reported elsewhere and not in the time period of interest	01 - Null entries 02 - Duplicates 03 - Language of interest 04 - Abstract that is reported elsewhere
1 - Nature of the study	05 - Treatment expression analysis 06 - Related articles
2 - Study population	07 - Not animals 08 - Not mammals 09 - Not human beings
3 - Outcome	10 - No functional link 11 - Ecological
4 - Potential	12 - Structure Analysis 13 - Clinical analysis 14 - Evo-devo 15 - Post transcription
Cannot decide	

Annex C: Symptom classification

Category	*Symptoms*
GH axis	GH IGF-1 IFG-BP3
Thyroid hormone axis	TSH T4 T3
Prolactin	Prolactin
Pituitary abnormality:	Empty sella Hypoplastic anterior pituitary
Midline abnormality:	Facial crowding, Prominent forehead Low nasal bridge Deep set eyes Acromicria Micrognathia Hypertelorism Umbilical hernia Downturned mouth Averted nostrils High pitched voice
Metabolic:	Hypoglycemia, Jaundice Hyperbilirubinemia Constipation Hypotonia Myxedema
Delayed growth and development:	Delayed bone maturation Microphallus Small kidneys Polyhydramnios Failure-to-thrive Short stature Late dentition Macroglossia Calf muscle dystrophy Open fontanelle Small facial skull
Developmental delay	Psychomotor deficit Hearing loss

Annex D: Mutations and ethnicity

Mutations	*Inheritance*	*DNA binding region*	*Country*	*Location*	*Reference*
Q4X	Recessive	NO	Malaysia	Exon 1	(Baş et al., 2015a)
Exon 1-2 del	Recessive	YES	Turkey	Exon 1,2	(Baş et al., 2015a)
P14L	Dominant (-)	NO	Russia	Exon 1	(Fofanova et al., 1998)
P24L	Dominant (-)	NO	Japan	Exon 1	(Ohta K, Nobukuni Y, Mitsubuchi H, Fujimoto S, Matsuo N, Inagaki H, Endo F, 1992) (Bano et al., 2010)
P76L	Dominant	NO	French	Exon 3	(Sobrier et al., 2016b) (Sobrier et al., 2016a)
F135C	Recessive	NO	Tunisian	Exon 3	(Ohta K, Nobukuni Y, Mitsubuchi H, Fujimoto S, Matsuo N, Inagaki H, Endo F, 1992)
R143L	Recessive	YES	Australia, Japan	Exon 3	(Ohta K, Nobukuni Y, Mitsubuchi H, Fujimoto S, Matsuo N, Inagaki H, Endo F, 1992)(McLennan et al., 2003)
R143Q	Recessive	YES	Turkey	Exon 3	(Aykut et al., 2014)
R143X	Recessive	YES	India	Exon 3	(Blum et al., 2018b)
K145X	Recessive	NO	Italy	Exon 3	(Hashimoto et al., 2003)
V153F	Recessive	NO	Turkish	Exon 4	(Baş et al., 2015b)
A158P	Recessive	NO	Germany, Dutch	Exon 4	(Pfäffle et al., 1992)
Q167K	Dominant	YES	Italy	Exon 4	(Malvagia et al., 2003)
Q167R	Dominant	YES	Russia	Exon 4	(Gavrilova et al., 2017)
R172Q	Recessive	YES	Malta, Japan	Exon 4	(J. P. G. Turton et al., 2005b)
R172X	Recessive	YES	Japan, Italian-America	Exon 4	(Brown et al., 1998)
E174G	Recessive	NO	Italian-American	Exon 4	(Brown et al., 1998)
S179R	Recessive	YES	Japan	Exon 4	(Miyata et al., 2006)
W193R	Recessive	NO	India	Exon 4	(Augustijn et al., 2001)
W193X	Recessive	NO	Israel	Exon 4	(G. Gat-Yablonski et al., 2005)
L194Q	Recessive	NO	Australia	Exon 4	(McLennan et al., 2003)
K216E	Dominant	NO	USA	Exon 5	(L. E. Cohen, Zanger, et al., 1999)

E230K	Recessive	NO	Malta, Russia, Israeli-Arab	Exon 6	(J. P. G. Turton et al., 2005a)(J. P. G. Turton et al., 2005b)
F233L	Recessive	NO	South East Asia	Exon 6	(Rainbow et al., 2005)
F233S	Recessive	NO	Taiwan	Exon 6	(Lee et al., 2011)
P239S	Recessive	NO	South West Saudi Arabia	Exon 6	(Pernasetti et al., 1998)
I244S	Recessive	NO	Turkish	Exon 6	(Baş et al., 2015a)(Baş et al., 2018)
E250X	Recessive	NO	Thailand	Exon 6	(Yoshifumi Irie et al., 1995)
F262L	Recessive	NO	Israel	Exon 6	(G. Gat-Yablonski et al., 2005)
R265W	Dominant	NO	Philippines, Turkey	Exon 6	(Bircan et al., 2001) (J. P. Turton et al., 2012)
R271W	Dominant	NO	United Kingdom, Russia, Australia, USA, Japan, Dutch, German, Swiss, Belgian, Norway, Canada	Exon 6	(Taha et al., 2005) (Ohta K, Nobukuni Y, Mitsubuchi H, Fujimoto S, Matsuo N, Inagaki H, Endo F, 1992) (Pine-Twaddell et al., 2013) (McLennan et al., 2003) (Rainbow et al., 2005)(J. P. G. Turton et al., 2005b)
V272X	Recessive	NO	Turkey	Exon 6	(Paper, 2002)
IVS1A ds +3 A-G	NA	NO	Philippines	Intron 1	(J. P. Turton et al., 2012)
IVS1 as -492 G-A	NA	NA	nd	NA	(Sun et al., 2006)
IVS2 ds +1 G-T	Recessive	NO	Japan	Exon 1	(Inoue et al., 2012)
IVS4 ds +1 G-A	Recessive	NO	Thai	Exon 4	(Snabboon et al., 2007)
snp-7057	NA	NO	China	3' UTR	(Sun et al., 2006)
rs300977	NA	NO	China	3' UTR	(Sun et al., 2006)
rs300996	NA	NO	China	3' UTR	(Sun et al., 2006)
New variant (−1295C > T)	NA	NO	Dutch	3' UTR	(Elizabeth et al., 2018)
rs10511134 (T > A)	NA	NO	Dutch	3' UTR	(Elizabeth et al., 2018)
rs300982 (C > T)	NA	NO	Dutch	3' UTR	(Elizabeth et al., 2018)
Q242R	Recessive	No	Ashkenazi Jews	Exon 6	(G. Gat-Yablonski et al., 2005)

E251NX	unk (hetero to W193R)	No	India	Exon 6	(Hendriks-Stegeman et al., 2001)
IVS2-3insA	Recessive	No	Italian	Exon 2	(Carlomagno et al., 2009)
c.502insT //p.Thr168IlefsX7	Recessive	Yes	Israel	Exon 4	(Tenenbaum-Rakover et al., 2011)
ins778A truncated at 284	unknown	No	Russia	Exon 6	(J. P. G. Turton et al., 2005a)
c.605-1G > A mutation - changes splicing	Recessive	No	India	Exon 4-5	(Birla et al., 2016)
c.605delC - premature stop at 220	Recessive	No	India	Exon 5	(Birla et al., 2016)
c.1-59T > A	Dominant	No	India	5' UTR	(Birla et al., 2016)
c. + 8C > T	Dominant	No	India	3' UTR	(Birla et al., 2016)
POU1F1, CHMP2B & VGLL3	Recessive	Yes	Israeli-Arab	Exon 1-6	(Galia Gat-Yablonski et al., 2011)

Annex E: Mutations pathology

Mutation	GH	IGF-1	IFG-BP3	**GH AXIS**	TSH	T4	T3	**Thyroid Hormone Axis**	PRL	**Prolactin**	Empty sella	Hypoplastic anterior	**Pituitary abnormality**	Facial crowding	Prominent forehead	Low nasal bridge	Deep set eyes	Acromicria	Micrognathia	Hypertelorism	Umbilical hernia	Downturned mouth	Anverted nostrils	High pitched voice	**Midline abnormality**	Hypoglicemia	Jaundice	Hyperbilirrubinemia	Constipation	Hypotone	Myxedema	**Metabolic**	Underdeveloped bones	Microphallus	Small kidneys	Polyhydramnios	Failure to thrive	Short stature	Late dentition	Macroglossia	Calf muscle distrophy	Open fontanelle	Small facial skull	**Delayed growth and**	Psychomotor deficit	Hearing loss	**Developmental delay**
WT	0	0	0	**0**	0	0	0	**0**	0	**0**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	**0**	0	0	**0**
P14L	1	0	0	**1**	1	0	0	**1**	1	**1**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
P24L	1	0	0	**1**	1	1	1	**3**	1	**1**	0	0	**0**	0	1	0	0	1	1	0	0	0	0	0	**3**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
P76L	1	0	0	**1**	0	0	0	**0**	0	**0**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	**0**	0	0	**0**
F135C	1	1	0	**2**	1	0	1	**2**	1	**1**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
R143L	1	0	0	**1**	1	0	0	**1**	1	**1**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
R143Q	1	1	1	**3**	1	1	0	**2**	1	**1**	0	0	**0**	0	0	0	0	0	0	0	0	1	1	0	**2**	0	0	0	0	1	0	**1**	0	0	0	0	0	0	0	1	0	0	**1**	0	0	**0**
V153F	1	0	0	**1**	1	0	0	**1**	1	**1**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	**0**	0	0	**0**
A158P	1	0	0	**1**	0	1	0	**1**	1	**1**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
Q167K	1	0	0	**1**	1	0	0	**1**	1	**1**	0	0	**0**	0	1	1	0	0	0	0	0	0	1	0	**3**	1	0	0	0	1	0	**2**	0	0	0	0	0	0	0	0	0	1	**1**	0	0	**0**
Q167R	1	1	0	**2**	0	1	0	**1**	1	**1**	0	0	**0**	0	1	0	0	0	0	0	0	0	0	0	**1**	1	0	0	0	0	0	**1**	0	0	0	0	1	0	0	0	0	0	**1**	1	0	**1**
R172Q	0	0	0	**0**	0	0	0	**0**	0	**0**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
E174G	1	0	0	**1**	1	1	0	**2**	1	**1**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	1	0	0	0	**1**	0	0	0	0	1	0	0	0	1	0	**2**	0	0	**0**
S179R	1	0	0	**1**	1	1	1	**3**	1	**1**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	1	0	0	0	**1**	0	0	0	0	1	1	0	0	0	0	**2**	0	0	**0**
W193R	1	1	1	**3**	1	1	0	**2**	1	**1**	0	0	**0**	0	0	1	0	0	0	0	0	0	0	0	**1**	0	0	0	1	1	1	**3**	0	0	0	0	0	1	0	1	0	1	**3**	0	0	**0**
L194Q	0	0	0	**0**	1	1	1	**3**	0	**0**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	1	0	0	1	0	**2**	0	0	0	0	1	1	0	0	0	0	**2**	0	0	**0**

K216E	1	0	0	**1**	1	1	0	**2**	1	**1**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	1	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	**0**
E230K	1	0	0	**1**	1	1	0	**2**	1	**1**	0	0	**0**	0	0	0	0	1	0	0	0	0	0	1	**2**	0	0	0	0	0	0	**0**	1	0	0	0	0	1	1	0	0	0	0	**3**	0	0	**0**
F233L	1	0	0	**1**	1	1	0	**2**	0	**0**	0	0	**0**	1	0	0	0	0	0	1	0	0	0	0	**2**	0	1	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	**0**
F233S	1	0	0	**1**	1	0	1	**2**	1	**1**	1	0	**1**	0	1	1	1	0	0	0	0	1	0	0	**4**	0	0	0	0	1	0	**1**	1	0	1	0	1	0	1	0	1	1	1	**7**	0	0	**0**
P239S	1	0	0	**1**	1	0	0	**1**	1	**1**	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
I244S	1	0	0	**1**	1	0	0	**1**	1	**1**	0	1	**1**	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	1	0	0	0	**1**	0	0	0	0	0	1	0	0	0	0	0	**1**	0	0	**0**
F262L	1	0	0	**1**	1	1	0	**2**	1	**1**	0	0	**0**	0	1	1	0	1	1	0	0	0	0	0	**4**	0	0	0	0	0	0	**0**	0	0	0	0	1	0	1	0	0	1	0	**3**	0	0	**0**
R265W	1	0	0	**1**	1	1	1	**3**	1	**1**	1	0	**1**	0	1	1	0	0	0	0	1	0	0	0	**3**	0	1	0	0	0	0	**1**	1	1	0	0	0	1	0	0	0	1	0	**4**	0	0	**0**
R271W	1	0	0	**1**	1	1	0	**2**	1	**1**	0	0	**0**	0	1	1	0	0	0	0	0	0	1	0	**3**	0	0	0	0	1	0	**1**	1	0	0	1	0	1	0	0	0	0	0	**3**	0	1	**1**

Annex F: Alignments and clashes

Mutation	C- score	Alignment score	Clashes	DNA binding energy
P14L	0.46	0.98708	0	-
P24L	0.58	0.98694	0	-
P76L	0.56	0.98843	0	-
F135C	0.6	0.98841	0	0.577721
R143L	0.47	0.98837	0	1.11551
R143Q	0.6	0.51022	0	0.798151
V153F	0.5	0.97957	0	0.655368
A158P	0.34	0.98912	0	0.601301
Q167K	0.52	0.98637	0	-0.55777
Q167R	0.51	0.98696	0	0.323343
R172Q	0.63	0.99247	0	0.772532
E174G	0.68	0.99148	0	0.676297
S179R	0.42	0.98833	0	-0.04331
W193R	0.66	0.98913	0	0.35194
L194Q	0.33	0.98748	0	0.579904
K216E	0.32	0.99068	0	0.811989
E230K	0.41	0.98854	0	0.276989
F233L	0.63	0.98793	0	0.653249
F233S	0.46	0.98888	0	0.574472
P239S	0.4	0.98441	0	0.47623
I244S	0.52	0.98729	0	0.424598
F262L	0.46	0.99207	0	0.39348
R265W	0.56	0.98763	0	0.942712
R271W	0.25	0.98724	0	0.814868

Annex G: Protein models

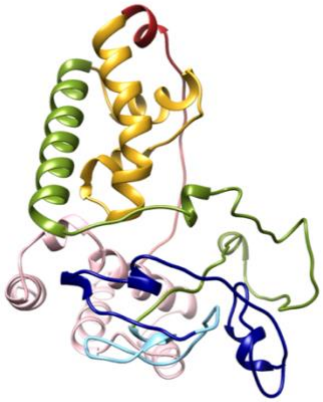


Figure 5A Exons 1

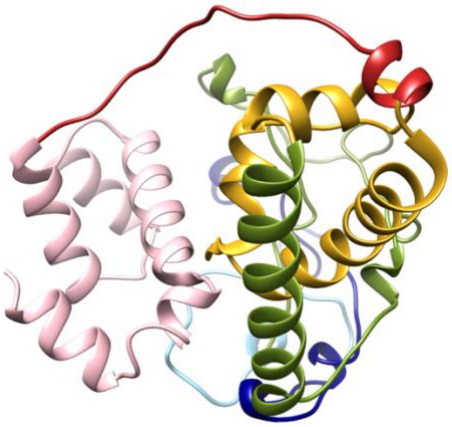


Figure 5B Exons 2

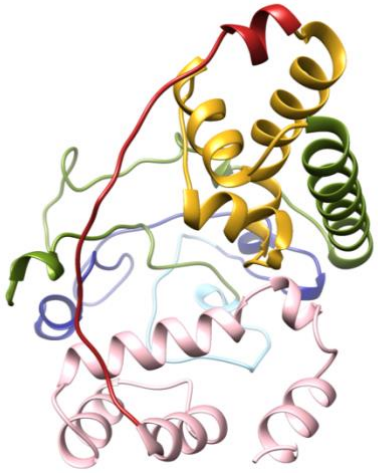


Figure 5C Exons 3

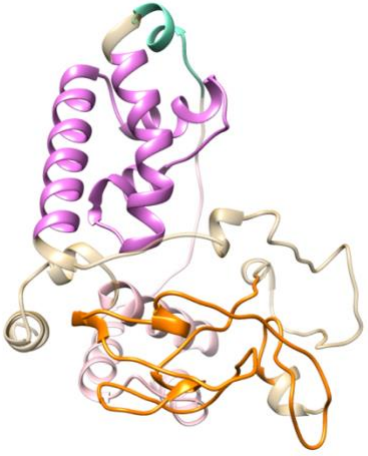


Figure 5D Domains 1

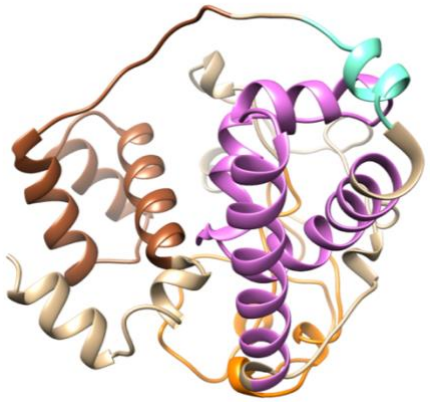


Figure 5E Domains 2

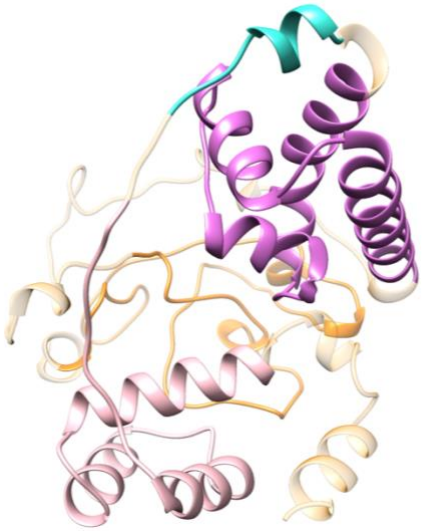
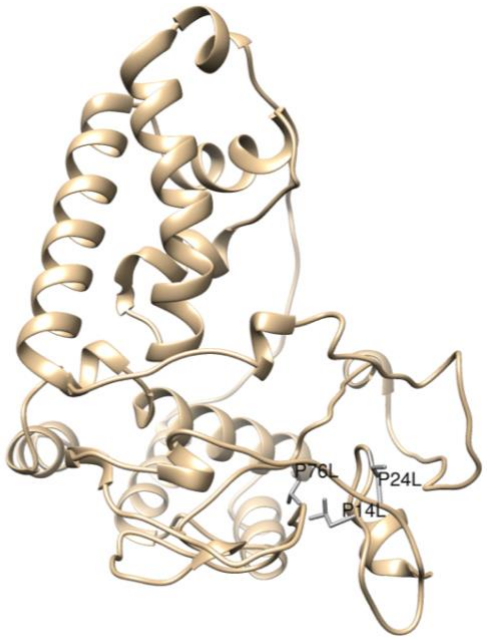
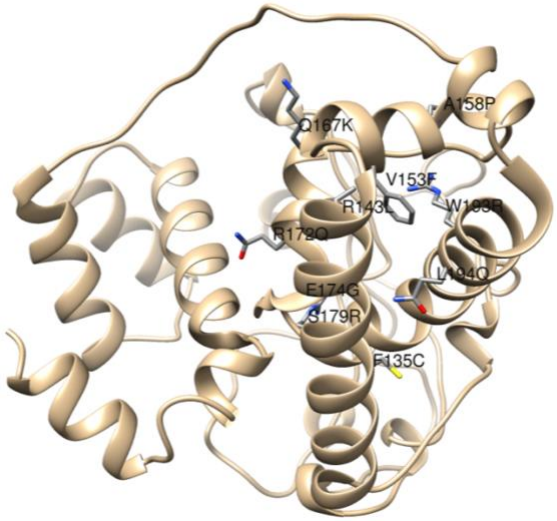


Figure 5F Domains 3





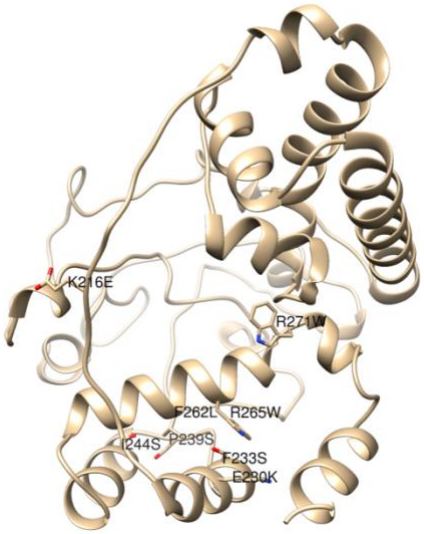


Figure 5G Mutations 1

Figure 5H Mutations 2

Figure 5I Mutations 3

POUF1: exons, domains and missense mutations

This protein structure was obtained through modeling with I-TASSER Model 4 obtained present a C-score of -3.70. Figures 5A to 5C present the six exons that code for POU1F1. Exon 1 in blue (AA 1-47), exon 2 in sky-blue (AA 48-71), exon 3 in olive green (AA 72 – 146), exon 4 in gold (AA 147 - 201), exon 5 in red (AA 202 – 222), and exon 6 in pink (AA 223 – 291). Figure 5D to 5F depict the POU1F1 domains as presented in Figure 3. The rest of the structure that does not fit into the functional domains is shown in cream color. Figure 5G shows the 3 mutations found in the TAD region, Figure 5H displays 10 mutations in the POUs two mutations: R143Q and Q167R are not shown in the figure, however mutations R143L and Q167K represent the position where these would occur, meaning in that in total there are 12 mutations described in this domain. Figure 5I exhibits 7 mutations occurring in the POUHD and 1 mutation in the carboxi terminal (R271W). Mutation F233L is not represented here, instead mutation F233S has been displayed.

Annex H: Dali scores