The American Journal of Human Genetics, Volume *106*

Supplemental Data

Bi-allelic Variants in *TKFC* Encoding Triokinase/FMN Cyclase Are Associated with Cataracts and Multisystem Disease

Saskia B. Wortmann, Brigitte Meunier, Lamia Mestek-Boukhibar, Florence van den Broek, Elaina M. Maldonado, Emma Clement, Daniel Weghuber, Johannes Spenger, Zdenek Jaros, Fatma Taha, Wyatt W. Yue, Simon J. Heales, James E. Davison, Johannes A. Mayr, and Shamima Rahman

SUPPLEMENTARY INFORMATION

Supplementary Figure 1:

Sequence comparison between human TKFC and yeast DAK1 and DAK2

% similarity human TKFC and DAK1:36%; human TKFC and DAK2:37%

```
Dak1           -MSAKSFEVTD-PVNSSLKGFALANPSITLVPEEKILFRKTD------SDKIALISGGGS 52
Dak2           -MSHKQFKSDGNIVTPYLLGLARSNPGLTVIKHDRVVFRTASAPNSGNPPKVSLVSGGGS 59
human-Dak      MTSKKLVNSVAGCADDALAGLVACNPNLQLLQGHRVALRSDLDS---LKGRVALLSGGGS 57
                 * * .:     .   * *:. .**.: ::  .:: :*.          :::*:*****

Dak1           GHEPTHAGFIGKGMLSGAVVGEIFASPSTKQILNAIRLVNE-NASGVLLIVKNYTGDVLH 111
Dak2           GHEPTHAGFVGEGALDAIAAGAIFASPSTKQIYSAIKAVE--SPKGTLIIVKNYTGDIIH 117
human-Dak      GHEPAHAGFIGKGMLTGVIAGAVFTSPAVGSILAAIRAVAQAGTVGTLLIVKNYTGDRLN 117
               ****:****:*:* * .  .* :*:**:. .*  **: *   .  *.*:******** ::

Dak1           FGLSAERARALGINCRVAVIGDDVAVGREKGGMVGRRALAGTVLVHKIVGAFAEEYSSKY 171
Dak2           FGLAAERAKAAGMKVELVAVGDDVSVGKKKGSLVGRRGLGATVLVHKIAGAAASH---GL 174
human-Dak      FGLAREQARAEGIPVEMVVIGDDSAFTVLK--KAGRRGLCGTVLIHKVAGALAEA---GV 172
               ***: *:*:* *:  .:..:*** :.   *   .***.* .***:**:.** *.

Dak1           GLDGTAKVAKIINDNLVTIGSSLDHCKVPGRKFESELNEKQMELGMGIHNEPGVKVLDPI 231
Dak2           ELAEVAEVAQSVVDNSVTIAASLDHCTVPGHKPEAILGENEYEIGMGIHNESGTYKSSPL 234
human-Dak      GLEEIAKQVNVVAKAMGTLGVSLSSCSVPGSKPTFELSADEVELGLGIHGEAGVRRIKMA 232
                *   *: .: : .   *:. **. *.*** *    *. .: *:*:***.* *.   .

Dak1           PSTEDLISKYMLPKLLDPNDKDRAFVKFDEDDEVVLLVNNLGGVSNFVISSITSKTTDFL 291
Dak2           PSISELVSQ-MLPLLL-DEDEDRSYVKFEPKEDVVLMVNNMGGMSNLELGYAAEVISEQL 292
human-Dak      T--ADEIVKLMLDHMT--NTTNASHVPVQPGSSVVMMVNNLGGLSFLELGIIADATVRSL 288
                   : : : **  :   :  : :.* .:  ..**::***:**:* : :.  :.     *

Dak1           KENYNITPVQTIAGTLMTSFNGNGFSITLLNATKATKALQSDFEEIKSVLDLLNAFTNAP 351
Dak2           IDKYQIVPKRTITGAFITALNGPGFGITLMNASKAGGDILKYFDYPTTASGWNQMYHSAK 352
human-Dak      E-GRGVKIARALVGTFMSALEMPGISLTLLLVDEP---LLKLIDAETTAAAWPNVAAV-- 342
                    :   :::.*:::::::  *:.:**: . :    : . ::  .:.    :

Dak1           GWPIADFEKT-SAPSVNDDLLHNEVTAKAVGTYDFDKFAEWMKSGAEQVIKSEPHITELD 410
Dak2           DWEVLAKGQVPTAPSLK--TLRNEK--GSGVKADYDTFAKILLAGIAKINEVEPKVTWYD 408
human-Dak      --SITGRKRSRVAPAEPQEAPDS----TAAGGSASKRMALVLERVCSTLLGLEEHLNALD 396
                  :    :   **:       .     :      .:*   :      :   * ::.  *

Dak1           NQVGDGDCGYTLVAGVKGITENLDKLS--KDSLSQAVAQISDFIEGSMGGTSGGLYSILL 468
Dak2           TIAGDGDCGTTLVSGGEALEEAIKNHTLRLEDAALGIEDIAYMVEDSMGGTSGGLYSIYL 468
human-Dak      RAAGDGDCGTTHSRAARAIQEWLKEGPP-PASPAQLLSKLSVLLLEKMGGSSGALYGLFL 455   G445S
                 .****** *   . ..: * :.:      . :  .:: ::  .***:**.**.: *

Dak1           SGFSHGLIQVCKSKDEPVTKEIVAKSLGIALDTLYKYTKARKGSSTMIDALEPFVKEFTA 528
Dak2           SALAQGVRD---SGDKELTAETFKKASNVALDALYKYTRARPGYRTLIDALQPFVEALKA 525
human-Dak      TAAAQPLKAK-------TSLPAWSAAMDAGLEAMQKYGKAAPGDRTMLDSLWAAGQELQA 508
               :. :: :           :      : . .*::: ** :*  *  *::*:*    : : *

Dak1           SKDFN--------KAVKAAEEGAKSTATFEAKFGRASYVGDS--------SQVEDPGAVG 572
Dak2           GKGPR--------AAAQAAYDGAEKTRKMDALVGRASYVAKEELRKLDSEGGLPDPGAVG 577
human-Dak      WKSPGADLLQVLTKAVKSAEAAAEATKNMEAGAGRASYISSARL-------EQPDPGAVA 561   R543I
                *.           *.::*  .*: * .::*  *****:..             *****.

Dak1           LCEFLKGVQSAL*--        584
Dak2           LAALLDGFVTAAGY*        591
human-Dak      AAAILRAILEVLQS-        575
```

Supplementary Table 1: Results of laboratory investigations

Parameter (reference range)	**Patient 1-1**	**Patient 1-2**	**Patient 2-1**	**Patient 2-2**
Full blood count	**Microcytic anaemia**	Normal	Normal	**Microcytic anaemia**
Vacuolated lymphocytes in peripheral blood film	Not seen	Not seen	ND	ND
Renal Function	Normal	Normal	Normal	Normal
Liver function: Albumin (35-52 g/L) ALT (10-25 U/L) Prothrombin time (8.2-14.1 seconds) INR (<1.2)	Normal 45 29 ND ND	**Abnormal:** **23** **900** 15.1-**23.5** ND	Normal	**Abnormal:** **23 - 38** **20- 386** ND < 1.2 -**1.4**
Blood lactate (<2 mmol/L)	1.2	**16**	Normal	0.9-**2.4**
Ammonia (<40 umol/L)	ND	25	ND	18
FGF21 (44 – 1515 pg/mL)	20	**4350**	ND	ND
Blood carnitine profile	Normal	Normal	ND	Normal
Plasma amino acids (Phenylalanine 21-93 µmol/l)	Normal profile	Normal profile	Mildly elevated phenylalanine	Phenylalanine 376-404
Very long chain fatty acids	Normal profile	Normal profile	ND	ND
Phytanate and pristanate	Normal	Normal	ND	ND
Transferrin electrophoresis	ND	Normal glycoforms	ND	ND
Cholesterol (112-189 mg/dl)	Normal	Normal	ND	Normal 91-147
Triglycerides (29-102 mg/dl)	Normal	Normal	ND	Elevated 58-**604**
Creatine kinase (75-230 U/L)	86	125	ND	ND
Urine organic acids	Elevated pyruvate, 3-hydroxbutyrate and acetoacetate	Mildly raised pyruvate (and dopamine metabolites reflecting inotrope therapy)	ND	3-methylglutaconic aciduria and elevated lactate excretion on one occasion, but normal at other times
Urine reducing substances	Negative	ND	ND	ND
Urine galactitol (3-17 mmol/mol creatinine)	**31**	ND	ND	ND

Urine lactose (0-34 mmol/mol creatinine)	**61**	ND	ND	ND
Other urine polyols	Normal	ND	ND	ND
Urine glycosaminoglycans	Not elevated	ND	ND	ND
Urine/plasma guanidino-acetate and creatine	Normal	ND	ND	ND
Red cell galactose-1-phosphate (<0.10 umol/gHb)	<0.10	ND	ND	ND
Galactose-1-phosphate uridyltransferase (15-35 umol/h/g Hb)	23.3	ND	ND	ND
Galactokinase (0.9-2.2 mol/hr/g Hb)	2.2	ND	ND	ND
Biotinidase (3.9 - 18.9 nmol/mL/min)		5.9	ND	ND
Bloodspot acid maltase activity	ND	Normal	ND	ND
Viral serology	ND	Negative	ND	ND

Key: ALT Alanine aminotransferase, FGF21 fibroblast growth factor 21

Supplementary methods

Whole genome sequencing in Family 1

1. **Genomic DNA (gDNA) preparation:** Whole genome sequencing (WGS) was performed on the trio of father, mother and affected individual P1-2. gDNA was extracted from blood samples in a diagnostic accredited lab (NE Thames Regional Genetics Lab).

   High quality gDNA was used for whole genome library preparation. 1ul gDNA was run on a 1% agarose gel to confirm absence of degradation. gDNA concentration was measured using Cubit dsDNA Broad Range Assay Kit (Invitrogen product #Q32850).

   DNA was diluted to 1.1ug in total volume of 55uL in HT1 buffer and transferred to Covaris 50uL individual tubes. gDNA was sheared to 350bp using E220 Focused-ultrasonicators (Covaris) for 60 seconds with the following parameters: target peak BP 400, Peak Incident Power 140, Duty factor 10% and 22 cycles of burst. Successful shearing was assessed on 1% agarose gel prior to starting library preparation.

2. **Library Preparation:** Whole genome gDNA libraries were prepared using TruSeq DNA PCR-Free Library Prep (Illumina FC-121-3001) following manufacturer advice starting with 1ug of sheared gDNA (in 50uL). Libraries were single indexed using Illumina's indexed adapters (Set A FC-121-3001 or Set B FC-121-3002). Library concentration was measured using quantitative PCR (qPCR) following the manufacturer's advice (KAPA Biosystems). Briefly, 2ul of library was diluted 10000x and 20000x in dilution buffer (100uL Tween, 2mL 1M Tris and 198mL dH_2O) and incubated overnight. qPCR was performed in triplicates in a total volume of 16uL each and run on an Applied Biosystem 7300 qPCR machine.

3. **Library normalization and sequencing:** Libraries were normalised to 2nM with Tris HCl (10mM) pH8.5 supplemented with 0.1% tween 20. Libraries were denatured with 0.2N NaOH and stabilised with 200mMTris HCl. gDNA libraries for the parents were pooled and sequenced on an Illumina NextSeq550 with a 2.7 pM loading concentration. The proband was sequenced on a double flow cell on an Illumina HiSeq Rapid Mode starting with 9pM loading concentration.

4. **Bioinformatics:** Read mapping and variant calling: Basecalling of raw sequencing reads was performed on BaseSpace Sequence hub (basespace.illumina.com). Fastq files for each individual were downloaded from BaseSpace, and reads from different lanes were merged together. Mapping and variant calling were performed using a GENALICE appliance running GENALICE Map 2.5.5 including Mapping, Variant Calling and the Population Calling module for trio analysis (GENALICE BV, Netherlands). GRCh build 37 and GENALICE default configuration files were used for WGS mapping, and trio variant detection. Aligned reads were stored in the GAR format (GENALICE Aligned Reads), using less than 5GB per sample. Variants were stored in a GVM (GENALICE Variant Map) per trio, using less than 200MB per sample. A standard multi sample VCF with Mendelian inheritance annotation using Context Based Call Enhancement was extracted from each GVM.

5. **Variant interpretation:**
   i. **Pre-filtering step**
      a. Common variants were filtered out (≤ 0.5% in 1000G[1], ExAC[2] and Exome Variant Server (evs.gs.washington.edu/EVS/) databases). For homozygous and hemizygous variants in proband, Minor Allele Frequency (MAF) was increased to ≤10% and variants with no homozygotes/ hemizygotes in ExAC were investigated.

b. Effect of variant on protein function was set to include predicted pathogenic, likely pathogenic and uncertain significance (benign or likely benign variants were investigated if further evidence of pathogenicity was available).
c. Variants associated with loss of function were kept, causing either: frameshift, in-frame in/del, missense or splice site (±7 nucleotides) alterations.
d. 5'UTR and 3'UTR variants were also investigated for genes known to be disease-causing or with compelling evidence for candidate genes.

ii. **Exclusion of known genetic causes:** This comprised setting a virtual gene panel as a filter to investigate genes associated with the patients' reported phenotypes as the first line of investigation. The gene panel was constructed by converting clinical phenotypes to HPO terms retrieving associated genes from different sources: The *Genomics England* PanelApp (https://panelapp.genomicsengland.co.uk/),Phenotips (https://phenotips.org/) OMIM Gene Map (https://www.omim.org/search/advanced/geneMap), established panels, UCL Great Ormond street Institute of Child Health disease experts and literature search in PubMed. A broader gene panel was analysed which consisted of variants in disease-associated genes from OMIM and DDG2P[3] databases.

6. **Allele Segregation**: allele segregation of the *TKFC* variant (c.1628G/T) was analysed in the family by PCR amplification followed by capillary sequencing (Sanger sequencing) using the following PCR primer set:

   TKFC_forward primer: 5'-TCCCTGCTGGAAGTAGATGAG-3'
   TKFC_reverse primer: 5'-CTGCAAGACCTCCAAGATGG-3'

Exome sequencing in Family 2

Exome sequencing was performed using the SureSelect Human All Exon 60 Mb Kit (Agilent) for enrichment and a HiSeq4000 (Illumina) for sequencing. The average coverage was 93-fold in P2-1, 138-fold in P2-2, 108-fold in the mother, and 126-fold in the father. The 20-fold coverage was 97.2% in P2-1, 99.0% in P2-2, 98.3% in the mother, and 98.3% in the father. We used BWA (version 0.5.8) for read alignment to the human reference assembly (hg19). Single-nucleotide variants (SNVs) and small insertions and deletions were detected with SAMtools as well as GATK. We excluded variants present with a frequency higher than 1% in 18000 control exomes in our database and public databases including gnomAD. Autosomal recessive, autosomal dominant, X-linked and inheritance via the mitochondrial DNA were considered.

Skin fibroblast culture and western blot in Family 1

Skin fibroblasts were subcultured in DMEM medium supplemented with 10% FBS, 10,000 U/mL penicillin-streptomycin, and 100 µg/mL uridine. Multiple cell passages were used as replicates for TKFC patient fibroblasts and multiple passages of fibroblasts from two metabolic disease control patients were used as controls. All cells were confirmed to be mycoplasma-free and collected at 90% confluency for protein extraction.

Protein was extracted with cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail in RIPA buffer for 30 minutes on ice. The lysed cell solution was centrifuged for 10 minutes at 16,000 xg and protein containing supernatant was collected. Protein concentration was measured by Pierce BCA assay kit as per the manufacturer's instructions.

Loading buffer (x6 solution) was added to samples prior to heat treatment for 10 mins at 70°C. Protein samples (30 µg of protein per lane) and ladder (Precision Plus Protein Kaleidoscope) were loaded into

Novex tris-glycine SDS 10% gels and SDS-PAGE ran at 120V for approximately 60 minutes in tris-glycine SDS running buffer.

Proteins were transferred onto Trans-Blot Turbo PVDF membrane via a Trans-Blot Turbo transfer system (Bio-Rad) at mixed molecular weight setting (1.3 A, 25 V for 7 minutes).

Blots were blocked with 0.1% skimmed milk-tris buffered saline-0.05% Tween 20 solution for 1 hour at room temperature. Primary antibodies were added to blocking solution (St Johns Laboratory Rabbit Anti-TKFC, STJ117616 at 1:1000; Proteintech Rabbit Anti-Vinculin; 26520-1-AP at 1:2000) and blots were probed overnight at 4°C. Membranes were washed for 5 minutes 3x prior to probing with secondary antibody (Cell Signaling Technology Goat Anti-Rabbit, 7074 at 1:2000 in blocking buffer) for 1 hour at room temperature. Blots were washed for 5 minutes x3 prior to visualisation by Thermo Scientific Pierce ECL Western blotting substrate (Vinculin after 5 second exposure; TKFC after 33 second exposure).

ImageJ 1.51j8 Java 1.8.0_112 (64-bit) was used to measure protein band density after converting images to 8-bit images; and GraphPad Prism v8.1.1 (330) for Windows 64-bit was used for graph generation and data analysis.

Preparation of recombinant TKFC protein

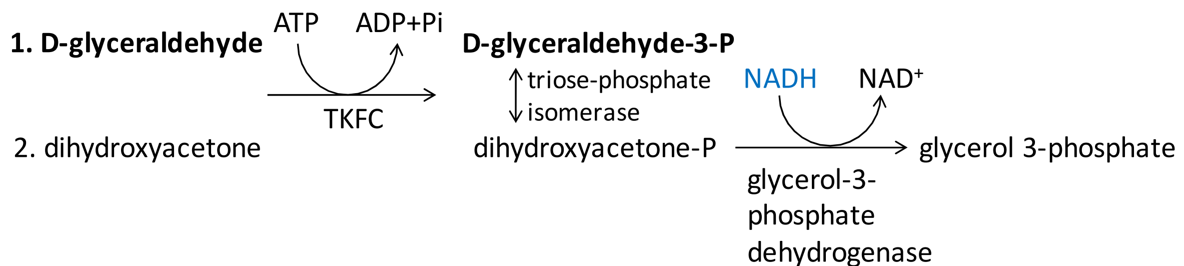
Wild-type human TKFC gene was cloned into the BamHI (…TAAGGATCCg**ATG**ACCTCC…) and BglII (…CAGAGC**TAA**AGATCTGCA…) sites of the pRSET-B vector providing an N-terminal 6xHis-tag. The two variants p.R543I and p.G445S were generated by site-directed mutagenesis (NEB-Builder, New England Biolabs) of the wild-type plasmid. The wild-type and mutated plasmids from several clones were Sanger sequenced in order to verify the mutated positions and to exclude any other variants. Wild-type or recombinant protein was expressed in *Escherichia coli* strain BL21(DE3)pLysS (Promega) grown on LB medium (yeast extract 5 g/l, tryptone 10 g/l, NaCl 5 g/l, pH 7.4) containing ampicillin (100 mg/l) and chloramphenicol (50 mg/l). Overnight cultures were inoculated into 100 ml of growth medium at a cell density of 0.1 optical density at 600 nm (OD600) and grown to an OD600 of 0.3 at 37°C under shaking at 200 rpm (revolutions per minute). The expression of recombinant TKFC was induced by the addition of 0.5 mmol/l IPTG (isopropyl β-D-1-thiogalactopyranoside) and shifting the temperature to 30 °C. The cells were harvested after three hours by centrifugation at 4500 rpm for 5 minutes. Cell pellets were washed with sterile water and resuspended in 6 ml of equilibration buffer (300 mmol/l NaCl, 10 mmol/l imidazole, 50 mmol/l sodium phosphate pH 7.4) and shock-frozen in liquid nitrogen. All following purification steps were performed either under ice-cooling or at 4°C (centrifugation steps).

After thawing, the cell suspensions were treated by ultrasonification with a micro sonification tip and sonification blast for 0.5 seconds and break for 2.5 seconds (Branson digital sonifier). Total sonification time, including brakes, was 5 minutes. After sonification the cell homogenates were centrifuged at 15000 x g for 20 minutes. The supernatants were loaded on HisPur cobalt spin columns (1 ml columns, Pierce Biotechnology)[4]. Spin columns had been washed twice with equilibration buffer before loading by centrifugation (700 x g, 2 minutes) to remove the column storage buffer. After the first 2 ml aliquot of the supernatant of the *E. coli* cell homogenates was loaded, the columns were closed on both ends, incubated for 10 minutes under ice cooling and shaking. Then they were centrifuged to separate the flow-through (700 x g, 2 minutes). This step was repeated since the volume of the cell homogenate exceeded the loading volume of the columns. After completing the loading step, the columns were washed three times with equilibration buffer. Finally, the His-tagged proteins were eluted with 3 x 1

ml of elution buffer (300 mmol/l NaCl, 150 mmol/l imidazole, 50 mmol/l sodium phosphate pH 7.4). The elution was performed by allowing the elution buffer to drip through (no centrifugation) and collecting the eluent in separate tubes. In order to remove the imidazole containing elution buffer, the eluent was diluted with dilution buffer (30 mmol/l NaCl, 30 mmol/l Tris pH 8.0) to a final volume of 15 ml, loaded on ultra centrifugal devices (Amicon Ultra 15, size exclusion 10 kDa) and centrifuged at 2500 x g for 25 minutes to reduce the volume to approximately 500 µl. This washing step was repeated twice and the remaining protein solution was stored in aliquots, which were shock frozen with liquid nitrogen and stored at -80 °C. Recombinant protein was quantified by polyacrylamide electrophoresis and equal amounts of wild-type and variant TKFC protein were used for enzymatic assays. At least two biological replicates were performed for wild-type and mutated proteins.

TKFC enzyme assay

The phosphorylation activity of TKFC was measured with either D-glyceraldehyde (GA) or dihydroxyacetone (DHA) as substrates in coupled assays, following the decrease of NADH at 340 nm by spectrophotometry in a 96 well plate reader at 37 °C for 20 minutes[5]:



The reaction mixture for GA phosphorylation contained a final concentration of 100 mmol/l Tris-HCl, pH 7.5, 0.18 mmol/l NADH, 10 mmol/l $MgCl_2$, 5 mmol/l ATP, 3 units/ml glycerol-3-phosphate dehydrogenase, 15 units/ml triose-phosphate isomerase and 0.1 mg/ml bovine serum albumin. Recombinant TKFC protein (25 µl of either wild-type, p.G455S or p.R543I) was added to the above mentioned mixture and pre-incubated for 5 minutes. The reaction was started by the addition of 12.5 µl of a 10 mmol/l D-glyceraldehyde (GA) stock solution to result in a final concentration of 0.5 mmol/l and final reaction volume of 250 µl.

For measuring the activity on dihydroxyacetone (DHA) the reaction was started by DHA instead of GA and did not contain triose-phosphate isomerase. Each measurement was performed in triplicate and from at least two replicates of recombinant protein purifications. The activity of the mutant proteins was related to wild-type protein, which was set to 100% activity.

Materials and Methods for Yeast Studies

Yeast strains

The control strain BY4742 (*α, his3, leu2, lys2, ura3*) and its derived isogenic deletion strains Δdak1 and Δdak2, were from Euroscarf (Frankfurt, Germany). The double mutant Δdak1Δdak2 was constructed by PCR-based deletion.

Overexpression of yeast DAK2 was obtained by transforming the control strain BY4742 and ∆dak mutants with a multi-copy plasmid (yEP352) containing DAK2 under the control of the PGK1 promoter.

Overexpression of human DAK was obtained by transforming the control strain BY4742 and ∆dak mutants with a yeast expression plasmid bearing human DAK under the control of the TEF1 promoter (from VectorBuilder, vectorbuilder.com)

Yeast growth media

The pre-culture medium contains 0.7% yeast nitrogen base, 2% glucose, 2% agar and 0.8 g/l of a complete supplement mixture with/without uracil or leucine supplied by Bio 101 (San Diego, CA, USA).

The DHA medium contains 0.7% yeast nitrogen base, 50 mM dihydroxyacetone (DHA), and 0.8 g/l of a complete supplement mixture with/without uracil supplied by Bio 101 (San Diego, CA, USA).

Supplementary References

1. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68-74.
2. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
3. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. American journal of human genetics 84, 524-533.
4. Banka, S., de Goede, C., Yue, W.W., Morris, A.A., von Bremen, B., Chandler, K.E., Feichtinger, R.G., Hart, C., Khan, N., Lunzer, V., et al. (2014). Expanding the clinical and molecular spectrum of thiamine pyrophosphokinase deficiency: a treatable neurological disorder caused by TPK1 mutations. Mol Genet Metab 113, 301-306.
5. Rodrigues, J.R., Couto, A., Cabezas, A., Pinto, R.M., Ribeiro, J.M., Canales, J., Costas, M.J., and Cameselle, J.C. (2014). Bifunctional homodimeric triokinase/FMN cyclase: contribution of protein domains to the activities of the human enzyme and molecular dynamics simulation of domain movements. The Journal of biological chemistry 289, 10620-10636.