

Comparative and Evolutionary Studies of Mitochondrial and Cytoplasmic Pyrophosphatase (*PPA*) Genes and Proteins

Roger S Holmes^{1,2*}, Kimberly D Spradling-Reeves¹ and Laura A Cox¹

¹Department of Genetics, Southwest National Primate Research Center, Texas Biomedical Research Institute, San Antonio, Texas, United States

²Eskitis Institute for Drug Discovery and School of Natural Sciences, Griffith University, Nathan, QLD, Australia

Abstract

Inorganic pyrophosphatase (*PPA*; PPase) (EC: 3.6.1.1) is a member of the diphosphatase enzyme family which functions as a diphosphate hydrolase within the cytoplasm (*PPA1*) and mitochondria (*PPA2*) of vertebrate tissues. *PPA1* and *PPA2* amino acid sequences and structures and *PPA*-like gene locations were examined using bioinformatic data from several genome projects. Sequence alignments and key conserved amino acid residues were also studied (human *PPA2* residues identified): the mitochondrial signal peptide (1-31); and active site residues responsible for Mg²⁺ binding (164Asp, 169Asp and 201Asp), substrate binding (127Arg) and serving as the proton donor site (138Tyr). Predicted 2D and 3D structures were identified for vertebrate *PPA1* and *PPA2* using the reported yeast *PPA1* structure (PDB: 1E9G). Vertebrate *PPA1* and *PPA2* genes usually contained 11 or 12 coding exons, respectively, with an extended exon 1 and an additional exon 3 observed for vertebrate *PPA2* genes. Transcription factor binding sites and CpG104 were identified within the human *PPA2* gene promoter; and MiR-590 for the *PPA2* 3'UTR. Phylogenetic analyses suggested that an ancestral invertebrate *PPA* gene underwent a gene duplication event to form 2 separate lines of vertebrate gene evolution: *PPA1* and *PPA2*.

Keywords: Vertebrates; Invertebrates; Amino acid sequence; Pyrophosphatase; Mitochondria; Cytoplasm; Evolution; Phylogeny; Phosphate metabolism

Abbreviations: *PPA*: inorganic pyrophosphatase; BLAST: Basic Local Alignment Search Tool; BLAT: Blast-Like Alignment Tool; NCBI: National Center for Biotechnology Information; UCSC: University of California Santa Cruz; KO: knock out; GI: gastro-intestinal; AceView: NCBI based representation of public mRNAs; SWISS-MODEL: automated protein structure homology-modeling server; TFBS: transcription factor binding sites; UTR: untranslated region

Introduction

Inorganic pyrophosphatases 1 and 2 (*PPA1*; *PPA2*; E.C.3.6.1.1; also designated as inorganic diphosphatase and diphosphate phosphohydrolase) are major diphosphate hydrolytic enzymes contained within the cytoplasm and mitochondria of vertebrate tissues and yeast cells [1-4]. These enzymes carry out essential roles in phosphate metabolism which ensure that biosynthetic reactions involving ATP hydrolysis and generating pyrophosphate are thermodynamically favored *in vivo* [5]. These include a wide range of biosynthetic reactions catalysed by ligases and synthetases [6]:

1. ATP+A+B=A-B+AMP+diphosphate [e.g., acetate CoA ligase EC: 6.2.1.1]
2. Diphosphate+H₂O → 2 Pi (inorganic phosphate) [*PPA* catalyzed hydrolysis EC: 3.6.1.1]

Structures for several *PPA1* and *PPA2* genes and cDNA sequences have been determined, including human (*Homo sapiens*) [4,7-8]; mouse (*Mus musculus*) [9]; rat (*Rattus norvegicus*) [10,11]; and cow (*Bos taurus*) [2] *PPA* genes. Human *PPA1*, which spans 31.1 kilobases and comprises 11 exons, is localized on chromosome 10; whereas human *PPA2* spans 302.7 kilobases and comprises 11 exons on chromosome 4 [4,12]. Both of these genes are widely expressed at very high levels in the body, consistent with their essential roles in catalyzing the hydrolysis of pyrophosphate to inorganic phosphate [12]. Cytoplasmic (*PPA1*) and mitochondrial (*PPA2*) forms of this enzyme have also been reported in yeast (*Saccharomyces cerevisiae*) with 50% sequence identities [3].

Using gene disruption techniques, Lundin and coworkers [1] have reported that both yeast *PPA1* and *PPA2* are required for cell growth with *PPA2* being essential for using respiratory carbon sources.

Three-dimensional structural analyses of yeast *PPA1* have shown that the enzyme belongs to the phosphoryl-transfer family with several key structural and catalytic features, including four divalent metal ions and a nucleophile hydroxide ion at the active site facilitating hydrolysis of the enzyme-bound inorganic pyrophosphate [13]. This paper reports the predicted gene structures and amino acid sequences for several vertebrate *PPA* genes and proteins, the predicted secondary and tertiary structures for vertebrate *PPA1* and *PPA2* protein subunits, and the structural, phylogenetic and evolutionary relationships for these genes and enzymes.

Materials and Methods

PPA gene and *PPA1/PPA2* protein identification

BLAST studies were undertaken using web tools from the NCBI (<http://www.ncbi.nlm.nih.gov>) [14]. Protein BLAST analyses used vertebrate *PPA1* and *PPA2* amino acid sequences previously described [4,7,10-11] (Table 1). Predicted *PPA*-like protein sequences were obtained in each case and subjected to protein and gene structure analyses.

BLAT analyses were undertaken for each of the predicted *PPA1* and *PPA2* amino acid sequences using the UCSC Genome

***Corresponding author:** Roger S Holmes, Eskitis Institute for Drug Discovery and School of Natural Sciences, Griffith University, Nathan, QLD, 4111 Australia, Tel: 61410583348; E-mail: r.holmes@griffith.edu.au

Received October 10, 2016; **Accepted** October 23, 2016; **Published** October 30, 2016

Citation: Holmes RS, Spradling-Reeves KD, Cox LA (2016) Comparative and Evolutionary Studies of Mitochondrial and Cytoplasmic Pyrophosphatase (*PPA*) Genes and Proteins. J Data Mining Genomics Proteomics 7: 205. doi: 10.4172/2153-0602.1000205

Copyright: © 2016 Holmes RS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Browser (<http://genome.ucsc.edu>) with the default settings to obtain the predicted locations for each of the vertebrate *PPA*-like genes, including exon boundary locations and gene sizes [15]. The structures for the major human *PPA1* and *PPA2* transcripts were obtained using the AceView website (<http://www.ncbi.nlm.nih.gov/ie/research/acembly/>) [12]. Alignments of *PPA* sequences with human *PPA1* and *PPA2* protein sequences were assembled using the Clustal Omega multiple sequence alignment program [16]. Predicted human *PPA1* and *PPA2* transcription factor binding sites (TFBS), micro-RNA binding sites (miR) and CpG islands [17] were examined using the UCSC Genome Browser [15].

Structures and predicted properties of *PPA1* and *PPA2* proteins

Predicted secondary and tertiary structures for human and other mammalian *PPA1* and *PPA2* proteins were obtained using the SWISS-MODEL web-server [18] and the reported tertiary structures for yeast *PPA1* [19] (PDB:1wgiA) with modeling residue ranges of 3-288 for human *PPA1* and 34-333 for human *PPA2* (Supplementary Fig 1). Identification of conserved domains for vertebrate *PPA1* and *PPA2* proteins was made using NCBI web tools [20].

Comparative human *PPA1* and *PPA2* gene expression

RNA-seq gene expression profiles across 53 selected tissues (or tissue segments) were examined from the public database for human *PPA1* and *PPA2*, based on expression levels for 175 individuals [21] (Data Source: GTEx Analysis Release V6p (dbGaP Accession phs000424.v6.p1) (<http://www.gtex.org>).

Phylogeny studies and sequence divergence

Phylogenetic analyses were undertaken using the <http://phylogeny.fr> platform [22]. Alignments of *PPA1* and *PPA2* sequences were assembled using MUSCLE [23] (Table 1). Alignment ambiguous regions were excluded prior to phylogenetic analysis yielding alignments for comparisons of these sequences. The phylogenetic tree was constructed using the maximum likelihood tree estimation program PhyML [24].

Results and Discussion

Alignments of *PPA1* and *PPA2* subunits

The deduced amino acid sequences for mouse and zebra fish (*Danio rerio*) *PPA1* and *PPA2* subunits are shown in Figure 1 together with the previously reported sequences for human [2,4,7] and yeast [1,3,13] *PPA1* and *PPA2* subunits (Table 1). Alignments of the human, mouse, zebrafish and yeast *PPA* subunits examined in this figure showed between 41% and 94% sequence identities, suggesting that these are products of two related families of genes and proteins, namely *PPA1* (cytosolic) and *PPA2* (mitochondrial) (Table 2). The amino acid sequences for vertebrate *PPA2* subunits contained 330-339 residues, whereas vertebrate *PPA1* subunits contained 289-291 residues (Figure 1; Table 1). These differences are predominantly explained by the presence of a mitochondrial leader peptide and an additional exon (exon 3) for the *PPA2* subunits. Table 1 summarizes this feature among all of the vertebrate *PPA1* and *PPA2* sequences examined and shows that vertebrate *PPA1* subunits had consistently lower pI values (5.2-5.9) as compared with vertebrate *PPA2* subunits (6.1-7.1) due to higher content of basic amino acid residues within the leader peptide and exon 3 sequences.

Several key amino acid residues or regions for vertebrate *PPA1* and *PPA2* have been recognized (sequence numbers refer to human

PPA2) (Figure 1). These include the mitochondrial N-terminus leader peptide (residues 1-31) [25]; metal binding residues at the active site (Mg^{2+}) (164Asp, 169Asp and 201Asp) [26]; N6-succinyllysine (216Lys, 259Lys), N6-acetyllysine (224Lys, 261Lys) and phosphoserine (317Ser) residues, which may have important cellular functions [27]; substrate binding site (127Arg); and proton donor site (138Tyr) (by analogy with yeast *PPA1*) [26].

Predicted secondary and tertiary structures of *PPA1* and *PPA2* subunits

Analyses of predicted secondary structures for vertebrate *PPA1* and *PPA2* sequences were compared with the previously reported structure for yeast *PPA1* [19] (PDB:1wgiA) (Figure 1). Similar α -helix and β -sheet structures were observed for the human *PPA1* and *PPA2* subunits examined, with 13 β -sheet and 3 α -helices predicted. Of particular interest was the clear separation of β -sheet and α -helix structures at the N-terminal and C-terminal ends of the *PPA* subunits, respectively. Consistent structures were particularly apparent near key residues or functional domains including the β -sheet and α -helix structures near the substrate binding site (127Arg), proton donor site (138Tyr) and the metal binding residues at the active site (Mg^{2+}) (164Asp, 169Asp and 201Asp) [26].

Figure 2 describes predicted tertiary structures for human *PPA1* and *PPA2* protein sequences, which showed significant similarities for these polypeptides with yeast *PPA1* [19]. Identification of specific structures within the predicted human *PPA1* and *PPA2* sequences was based on the reported structure for yeast *PPA1* [19]. The active sites for *PPA1* and *PPA2* were centrally located and surrounded by several β -sheet structures (β 1- β 9) as well as the metal binding residues at the active site (Mg^{2+}) (164Asp, 169Asp and 201Asp) [26]. These comparative studies of vertebrate and yeast *PPA* proteins suggest that the properties, structures and key sequences are substantially retained for all of the sequences examined, reflecting a highly conserved structure throughout invertebrate and vertebrate evolution.

Predicted gene locations and exonic structures for vertebrate and invertebrate *PPA* genes

Table 1 summarizes the predicted locations for vertebrate and invertebrate *PPA* genes based upon BLAT interrogations of genomes using the reported sequences for human, mouse, rat, bovine and yeast *PPA1* and *PPA2* [1-4,7,10,11,13,19,26] and the predicted sequences for other *PPA1* and *PPA2* proteins and the UCSC Genome Browser [15]. Human *PPA1* and *PPA2* genes were located on different chromosomes (chromosomes 10 and 4, respectively), which is observed for all vertebrate genomes examined, and for the yeast genome (Table 1). Of particular interest to the evolution of *PPA*-like genes in invertebrate genomes, three invertebrate genomes (sea squirt (*Ciona intestinalis*), sea urchin (*Strongylcentrus purpuratus*) and worm (*Caenorhabditis elegans*) showed evidence of having only one gene which was similar to the vertebrate *PPA2* gene in containing a mitochondrial leader peptide. This was in contrast to the yeast genome which contained *PPA1*- and *PPA2*-like single exon genes encoding cytosolic and mitochondrial enzymes, respectively [1,3].

Figure 1 summarizes the predicted exonic start sites for human, mouse and zebrafish *PPA1* and *PPA2* genes with each having 11 or 12 exons, respectively, in identical or similar positions. Of particular interest to this comparison were the extended exon 1 sequence (which encoded the mitochondrial leader peptide sequence) and the additional exon 3 'inserts' for the vertebrate *PPA2* genes and proteins. In contrast

Gene	Organism	Species	Chromosome [^] location	Coding Exons (strand)	Gene Size bps	GenBank ID*	UNIPROT ID	Amino acids	Subunit MW (pI)	Leader Peptide
PPA1	Human	<i>Homo sapiens</i>	10:70,204,874-70,233,327	11 (-ve)	28,454	BC105034	Q15181	289	32,660 (5.5)	NA
PPA1	Baboon	<i>Papio anubis</i>	9:59,513,291-59,539,827	11 (+ve)	26,537	XP_003903878*	A0A096P4X1	289	32,646 (5.5)	NA
PPA1	Mouse	<i>Mus musculus</i>	10:61,648,728-61,672,432	11 (+ve)	23,705	BC10468	Q9D819	289	32,667 (5.4)	NA
PPA1	Opossum	<i>Mondelphis domestica</i>	1:5,572,772-5,601,324	11 (-ve)	28,553	XP_007478020*	F7CY62	295	32,649 (5.9)	NA
PPA1	Chicken	<i>Gallus gallus</i>	6:11,078,718-11,085,692	11 (-ve)	6,975	XP_001232700*	F1NT28	290	32,637 (5.5)	NA
PPA1	Lizard	<i>Anolis carolinensis</i>	3:51,244,145-51,275,621	11 (+ve)	31,477	XP_003218578*	G1K9N5	287	32,181 (5.3)	NA
PPA1	Frog	<i>Xenopus tropicalis</i>	[^] GL172815:122,605-160,928	11 (+ve)	38,324	BC088578	F6PVD4	289	32,953 (5.3)	NA
PPA1	Zebra fish	<i>Danio rerio</i>	13:30,022,211-30,031,076	11 (-ve)	8,866	BC164460	Q568N9	291	32,644 (5.2)	NA
PPA2	Human	<i>Homo sapiens</i>	4:105,369,725-105,474,050	12 (-ve)	#####	BC022803	Q9H2U2	334	37,920 (7.1)	1..31
PPA2	Baboon	<i>Papio anubis</i>	5:96,645,703-96,733,406	12 (-ve)	87,704	XP_009205561*	A0A096MN84	334	38,001 (6.1)	1..31
PPA2	Mouse	<i>Mus musculus</i>	3:132,973,060-133,041,199	12 (+ve)	68,139	BC011417	Q91VM9	330	38,115 (6.5)	1..26
PPA2	Opossum	<i>Mondelphis domestica</i>	5:43,324,178-43,449,553	12 (+ve)	#####	XP_007495904*	F7EMS4	363	41,604 (6.7)	1..53
PPA2	Chicken	<i>Gallus gallus</i>	4:38,190,439-38,219,514	12 (+ve)	29,076	XP_004941113*	E1C6X1	317	35,662 (6.7)	1..31
PPA2	Lizard	<i>Anolis carolinensis</i>	[^] GL343375:367,173-392,999	12 (+ve)	25,827	XP_008117406*	H9GE26	355	40,043 (7.1)	1..48
PPA2	Frog	<i>Xenopus tropicalis</i>	[^] KB021649:148,818,480-148,835,143	12 (+ve)	16,664	CR761093	Q28H25	335	38,010 (6.5)	1..29
PPA2	Zebra fish	<i>Danio rerio</i>	1:25,698,526-25,704,534	12 (-ve)	6,009	XP_005159941*	A0A0R4IRJ2	339	38,433 (6.7)	1..38
PPA	Sea squirt	<i>CioNA intestiNalis</i>	13q:228,907-235,200	9 (+ve)	6,924	XP_002124420*	NA	329	37,041 (5.5)	1..40
PPA	Sea urchin	<i>Strongylocentrus purpuratus</i>	[^] 85383:969-13466	10 (-ve)	12,498	XP_793193*	W4YMK8	335	37,865 (6.3)	1..43
PYP1	Worm	<i>Caenorhabditis elegans</i>	2R:24,559,833-24,560,387	6 (-ve)	556	CAA93107	Q18680	427	46,330 (5.8)	1..109
PPA1	Yeast	<i>Saccharomyces cerevisiae</i>	II:257,112-257,975	1 (-ve)	863	DAA07132	P00817	287	32,300 (5.4)	NA
PPA2	Yeast	<i>Saccharomyces cerevisiae</i>	XIII:801,772-802,701	1 (+ve)	930	BK006946	P28239	310	35,573 (7.1)	1..42

*=predicted sequence; [^]=gene scaffold ID; pI=isoelectric point; bps=base pairs of nucleotide sequence; NA=not available.

Table 1: Vertebrate PPA1 and PPA2 and invertebrate PPA-like genes and subunits.

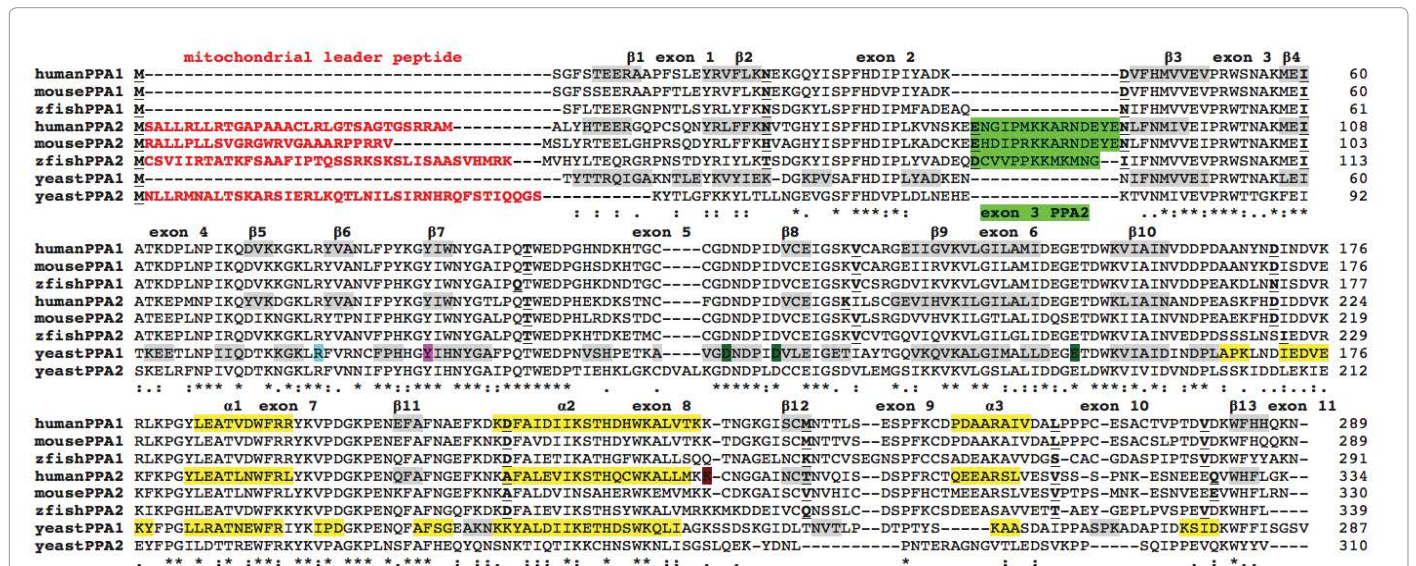


Figure 1: Amino acid sequence alignments for vertebrate and yeast PPA1 and PPA2 subunits. See Table 1 for sources of PPA1 and PPA2 sequences; * shows identical residues for PPA subunits; similar alternate residues; . dissimilar alternate residues; mitochondrial leader peptide residues are in red; helix (yeast PPA1 or predicted helix); sheet (yeast PPA1) or predicted sheet; active site Arg127 shown in blue; proton donor 138Tyr shown in pink; Mg²⁺ binding residues for yeast PPA1 are in dark green; bold font shows known or predicted exon junctions; exon numbers refer to human PPA1 gene; note an additional exon was observed for vertebrate PPA2 genes shown in light green; nsSNP variant for human PPA2 gene (rs13787 282K→N) shown in brown.

to vertebrate PPA2 genes and proteins, the yeast PPA2 gene and protein (a single exon gene) encoded only the mitochondrial leader peptide but was without the vertebrate exon 3 'insert' sequence.

Figure 3 illustrates the predicted structures of mRNAs for human PPA1 and PPA2 transcripts for the major transcript isoform in each case [12]. The transcripts were 31.1 and 30.3 kilobases in length,

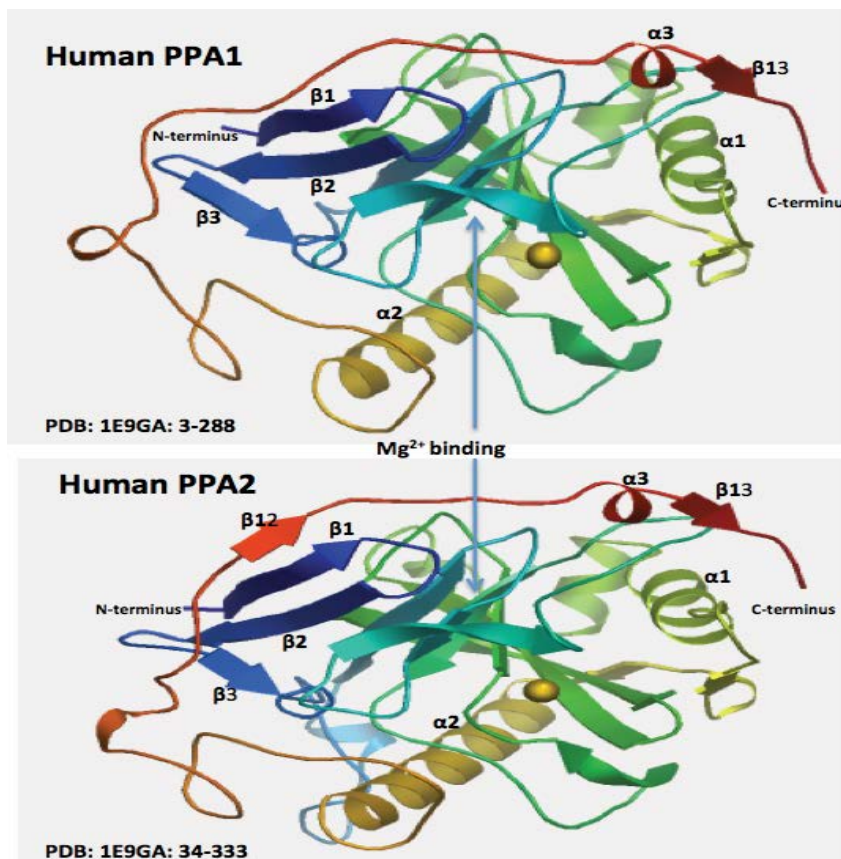


Figure 2: Predicted tertiary structures for human *PPA1* and *PPA2* subunits. The predicted human *PPA1* and *PPA2* 3D structures were obtained using the SWISS MODEL web site <http://swissmodel.expasy.org> and based on the reported structure for yeast *PPA1* (PDB: 1E9GA) [19]; the rainbow color code describes the 3D structures from the N- (blue) to C-termini (red color); N refers to amino-terminus; C refers to carboxyl terminus; specific α helices ($\alpha 1$, $\alpha 2$..) and β sheets ($\beta 1$... $\beta 13$) were identified, as well as the active site region and the Mg^{2+} binding sites.

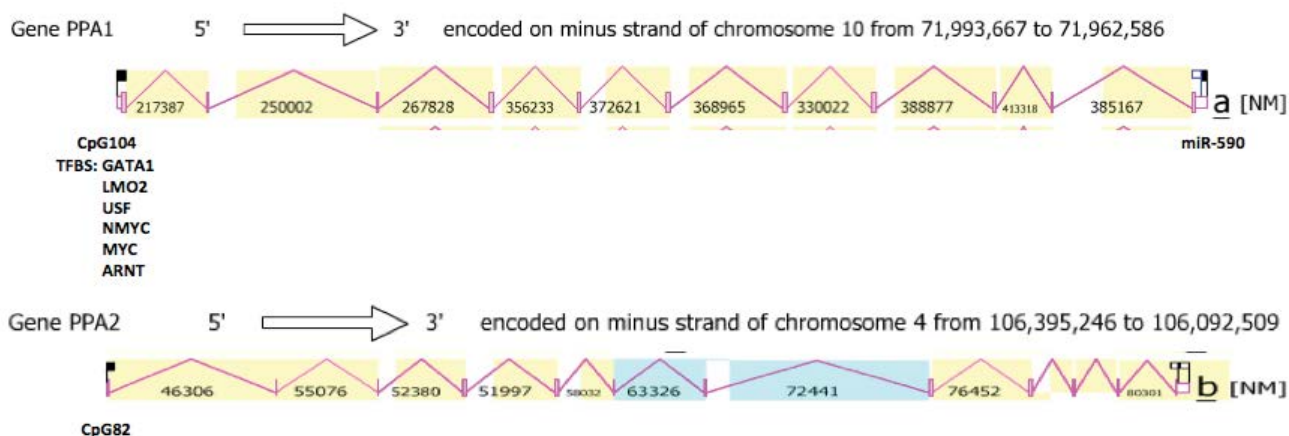


Figure 3: Gene Structures for the Human *PPA1* and *PPA2* genes. Derived from the AceView website <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/> [12]; the major isoform variant is shown with capped 5'- and 3'- ends for the predicted mRNA sequences; introns (pink lines) and exons (pink boxes) are shown; the length of the mRNAs (as kilobases or kb) are shown; CpG islands (CpG104 and CpG82) are shown for the *PPA1* and *PPA2* promoters, respectively; a miRNA590 binding site was identified for the 3'UTR of the human *PPA1* gene; the direction for transcription is shown; TFBS refers to transcription factor binding sites located within the *PPA1* gene promoter.

respectively, with 10 introns and 11 exons present for the *PPA1* mRNA transcript; and 11 introns and 12 exons for the *PPA2* mRNA transcript. The human *PPA1* gene promoter contained a CpG island (CpG104)

and six predicted TFBS: *GATA1*, a transcriptional activator or repressor which acts as a switch factor for erythroid development [28]; *LMO2*, which acts with *TALI/SCL* (master regulators of erythropoiesis) to

	Human PPA1	Mouse PPA1	Zebra fish PPA1	Human PPA2	Mouse PPA2	Zebra fish PPA2	Yeast PPA1	Yeast PPA2
Human PPA1	100	94	72	62	61	67	53	44
Mouse PPA1	94	100	72	60	60	64	54	45
Zebra fish PPA1	72	72	100	58	58	68	55	46
Human PPA2	62	60	58	100	77	59	51	43
Mouse PPA2	61	60	58	77	100	57	51	41
Zebra fish PPA2	67	64	68	59	57	100	53	46
Yeast PPA1	53	54	55	51	51	53	100	50
Yeast PPA2	44	45	46	43	41	46	50	100

Table 2: Percentage identities for vertebrate and yeast PPA1 and PPA2 subunit amino acid sequences. Sequences obtained from sources in Table 1.

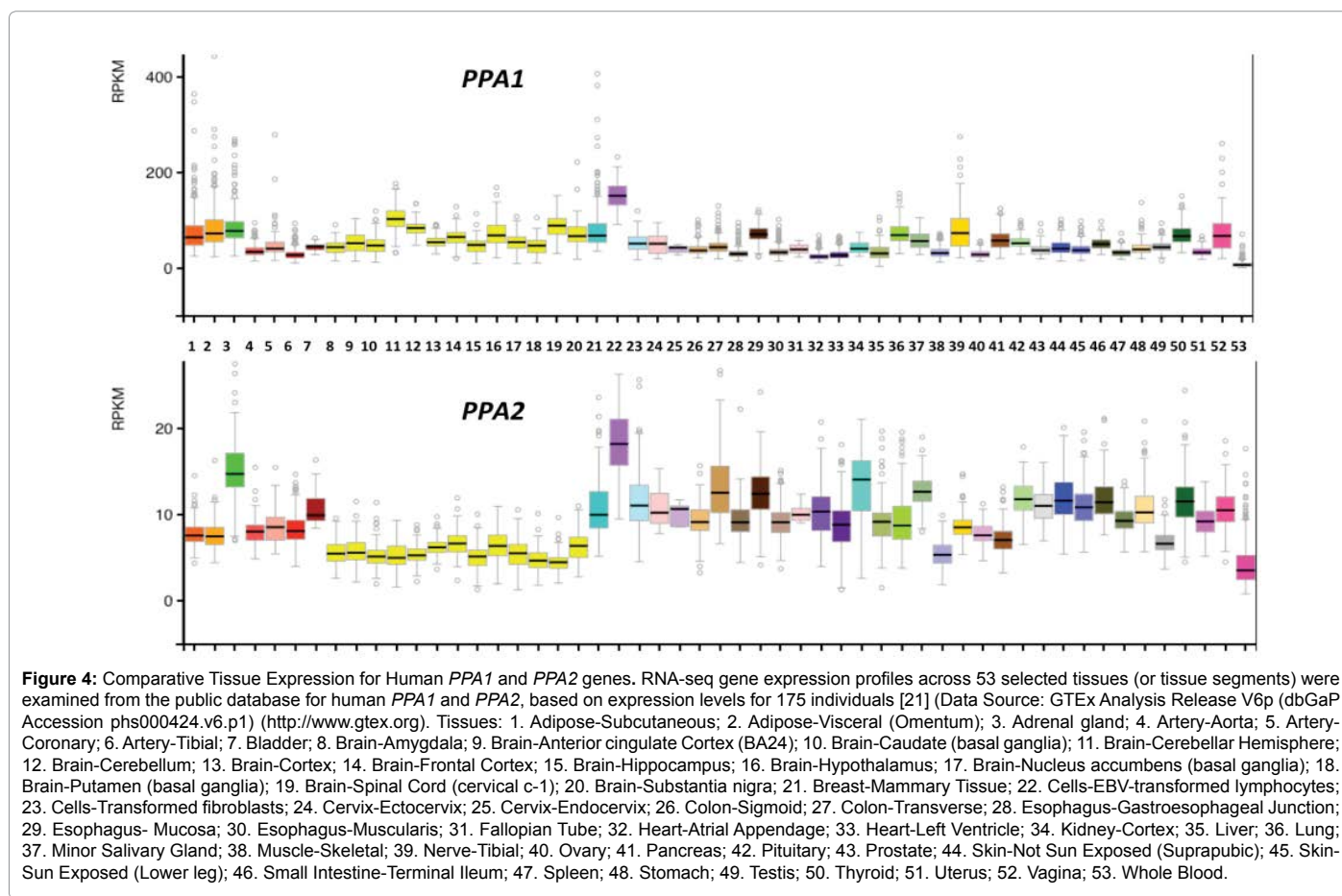


Figure 4: Comparative Tissue Expression for Human PPA1 and PPA2 genes. RNA-seq gene expression profiles across 53 selected tissues (or tissue segments) were examined from the public database for human PPA1 and PPA2, based on expression levels for 175 individuals [21] (Data Source: GTEx Analysis Release V6p (dbGaP Accession phs000424.v6.p1) (<http://www.gtex.org>). Tissues: 1. Adipose-Subcutaneous; 2. Adipose-Visceral (Omentum); 3. Adrenal gland; 4. Artery-Aorta; 5. Artery-Coronary; 6. Artery-Tibial; 7. Bladder; 8. Brain-Amygdala; 9. Brain-Anterior cingulate Cortex (BA24); 10. Brain-Caudate (basal ganglia); 11. Brain-Cerebellar Hemisphere; 12. Brain-Cerebellum; 13. Brain-Cortex; 14. Brain-Frontal Cortex; 15. Brain-Hippocampus; 16. Brain-Hypothalamus; 17. Brain-Nucleus accumbens (basal ganglia); 18. Brain-Putamen (basal ganglia); 19. Brain-Spinal Cord (cervical c-1); 20. Brain-Substantia nigra; 21. Breast-Mammary Tissue; 22. Cells-EBV-transformed lymphocytes; 23. Cells-Transformed fibroblasts; 24. Cervix-Ectocervix; 25. Cervix-Endocervix; 26. Colon-Sigmoid; 27. Colon-Transverse; 28. Esophagus-Gastroesophageal Junction; 29. Esophagus- Mucosa; 30. Esophagus-Muscularis; 31. Fallopian Tube; 32. Heart-Atrial Appendage; 33. Heart-Left Ventricle; 34. Kidney-Cortex; 35. Liver; 36. Lung; 37. Minor Salivary Gland; 38. Muscle-Skeletal; 39. Nerve-Tibial; 40. Ovary; 41. Pancreas; 42. Pituitary; 43. Prostate; 44. Skin-Not Sun Exposed (Suprapubic); 45. Skin-Sun Exposed (Lower leg); 46. Small Intestine-Terminal Ileum; 47. Spleen; 48. Stomach; 49. Testis; 50. Thyroid; 51. Uterus; 52. Vagina; 53. Whole Blood.

regulate erythroid cell development [29]; *USF* (upstream regulatory factor), a transcription factor controlling expression of multiple genes involved in lipid and glucose homeostasis [30]; *NMYC* and *MYC*, encoding N-myc and Myc proto-oncogene proteins, which support the development of the epidermis and central nervous system [31]; and *ARNT*, encoding the aryl hydrocarbon receptor nuclear translocator, which participates in the activation of aryl hydrocarbon procarcinogens and serves as an essential regulator of hematopoietic viability [32].

A microRNA site (miR-590) was also located in the 3'-UTR of human *PPA1*, which is potentially of major significance for the regulation of this gene (Figure 3). A recent study of miR-590 has shown that it regulates osteogenic differentiation in developing human mesenchymal cells [33]. Moreover, this comparative study of mammalian miR-590 nucleotide sequences, located within the 3'-UTR of mammalian *PPA1* genes, showed that the UAAAAUUA sequence is strictly conserved in all genes examined, with the exception of the bovine *PPA1* 3'UTR

sequence, which contained an additional nucleotide (UAAAAUUA). This high level of conservation among mammalian *PPA1* miR-590 sequences may reflect an important role in erythroid development and osteogenesis. Moreover, *PPA1* expression has been shown to induce type 1 collagen mRNA expression in osteoclasts, supporting a broader role of the *PPA1* gene in osteoblast development [34].

Comparative PPA1 and PPA2 human tissue expression

Figure 4 shows comparative gene expression for various human tissues obtained from RNA-seq gene expression profiles for human *PPA1* and *PPA2* genes obtained for 53 selected tissues or tissue segments for 175 individuals [21] (Data Source: GTEx Analysis Release V6p (dbGaP Accession phs000424.v6.p1) (<http://www.gtex.org>). These data supported a much higher level of tissue expression for human *PPA1*, particularly for lymphocytes, adrenal gland, adipose tissue and various regions of the brain. High *PPA1* gene expression levels have

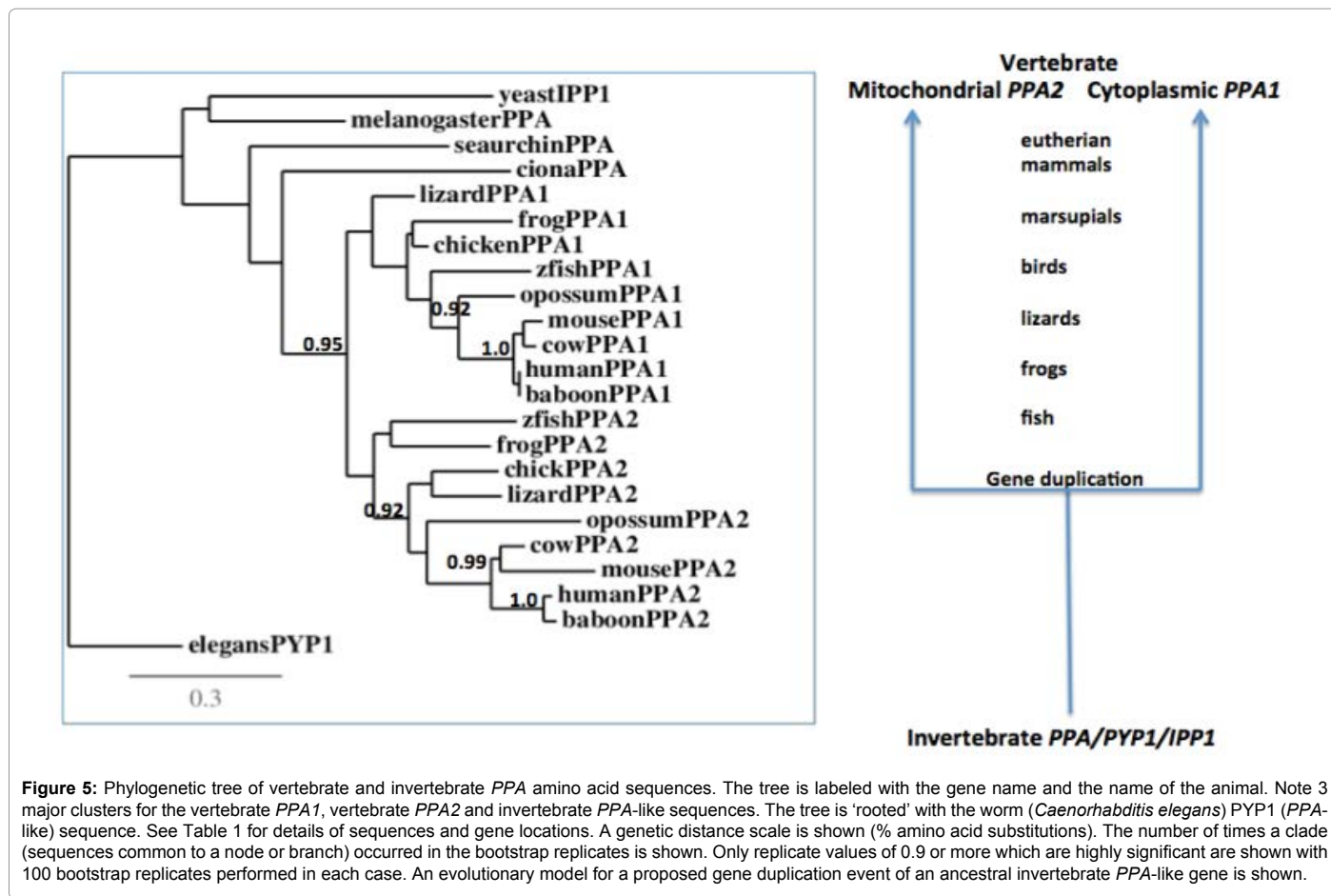


Figure 5: Phylogenetic tree of vertebrate and invertebrate PPA amino acid sequences. The tree is labeled with the gene name and the name of the animal. Note 3 major clusters for the vertebrate PPA1, vertebrate PPA2 and invertebrate PPA-like sequences. The tree is 'rooted' with the worm (*Caenorhabditis elegans*) PYP1 (PPA-like) sequence. See Table 1 for details of sequences and gene locations. A genetic distance scale is shown (% amino acid substitutions). The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates is shown. Only replicate values of 0.9 or more which are highly significant are shown with 100 bootstrap replicates performed in each case. An evolutionary model for a proposed gene duplication event of an ancestral invertebrate PPA-like gene is shown.

also been reported for mammalian spermatazoa where PPA1 may play a role in an inorganic pyrophosphate metabolizing, ATP substituting pathway [35]. PPA2 levels were, however, consistently 10 times lower for human tissues examined than for PPA1, but with higher levels of PPA2 expression observed in lymphocytes, adrenal gland, kidney cortex and mammary tissue. The presence of six TFBS within the PPA1 gene promoter (*GATA1*; *LMO2*; *USF*; *NMYC*, *MYC* and *ARNT*) and the absence of any predicted TFBS within the PPA2 gene promoter may contribute to this order of magnitude difference in expression level between PPA1 and PPA2.

Phylogeny and divergence of vertebrate PPA1 and PPA2

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of human and other vertebrate PPA1 and PPA2 amino acid sequences with invertebrate (yeast, fruit fly, sea urchin and sea squirt) sequences. The phylogram was 'rooted' with a *C. elegans* PYP1 sequence and showed clustering of the PPA-like sequences into several groups: invertebrate PPA-like sequences; vertebrate PPA1 (cytoplasmic) sequences; and vertebrate PPA2 (mitochondrial) sequences. Overall, these data suggest that the vertebrate PPA1 and PPA2 genes arose from a gene duplication event of an ancestral invertebrate PPA-like gene, resulting in at least two separate lines of vertebrate gene evolution for PPA1-like and PPA2-like genes. This is supported by the comparative biochemical and genomic evidence for vertebrate PPA1 and PPA2-like genes and encoded proteins, which shared several key features of protein and gene structure, including having similar α - β secondary and tertiary structures (Figure 1). In addition, the locations of vertebrate

PPA1 and PPA2 genes reported on separate chromosomes (Table 1) may reflect on a possible mechanism for ancestral vertebrate PPA gene duplication by whole-genome duplication rather than by an unequal crossover event of a single ancestral chromosome, as exemplified by studies of early vertebrate hemoglobin genes and proteins [36].

Conclusion

In conclusion, the results of the present study indicate that vertebrate PPA1 and PPA2 genes and encoded PPA (PPA1 and PPA2) enzymes represent a distinct diphosphate hydrolytic enzyme and gene family which share key conserved sequences and structures with those reported for yeast PPA1 and PPA2 gene families. PPA1 is a major cytoplasmic pyrophosphatase expressed at high level in many tissues of the body, particularly lung, kidney, testis, liver and ovary, where it contributes to driving biosynthetic reactions involving ATP hydrolysis and generating pyrophosphate, which is hydrolysed and thermodynamically favored *in vivo*. PPA2 is a major mitochondrial enzyme performing similar roles in driving biosynthetic reactions although expressed at somewhat lower levels but with a wide tissue expression profile, similar to that for PPA1. Bioinformatic methods were used to predict the amino acid sequences, secondary and tertiary structures and gene locations for PPA1 and PPA2 genes and encoded proteins using data from several vertebrate genome projects. Vertebrate PPA protein subunits shared 58-94% sequence identities and exhibited sequence alignments and identities for key PPA amino acid residues as well as extensive conservation of predicted secondary and tertiary structures with those previously reported for yeast PPA1. Phylogenetic

analyses demonstrated the relationships and potential evolutionary origins of the vertebrate *PPA1* and *PPA2* gene families which were related to a worm (*Caenorhabditis elegans*) pyrophosphatase (*PYP1*) gene and to yeast, fruit fly and sea urchin *PPA*-like genes and proteins. These studies indicated that *PPA1* and *PPA2* genes may have appeared early in vertebrate evolution following gene duplication of an ancestral *PPA*-like gene, following whole-genome duplication in the vertebrate ancestor.

Acknowledgement

Research reported in this manuscript was supported by National Institutes of Health (NIH) R01 HL118556. This investigation was conducted in facilities constructed with support from ORIP through grant numbers C06 RR14578, C06 RR15456, C06 RR013556, and C06 RR017515.

References

- Lundin M, Balatscheffsky H, Ronne H (1991) Yeast *PPA2* gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. J Biol Chem 266: 12168-12172.
- Yang Z, Wenzel TG (1992) Molecular cloning and functional expression of cDNA encoding a mammalian inorganic pyrophosphatase. J Biol Chem 267: 24641-24647.
- Vihinen M, Lundin M, Baltscheffsky H (1992) Computer modeling of two inorganic pyrophosphatases. Biochem Biophys Res Commun 186: 122-128.
- Curbo S, Lagier-Tourenne C, Carrozzo R, Palenzuela L, Luciola S, et al. (2006) Human mitochondrial pyrophosphatase: cDNA cloning and analysis of the gene in patients with mtDNA depletion syndromes. Genomics 87: 410-416.
- Kankare J, Salminen T, Lahti R, Cooperman BS, Baykov AA, et al. (1996) Structure of *Escherichia coli* inorganic pyrophosphatase at 2.2 Å resolution. Acta Crystallogr D Biol Crystallogr 52: 551-63.
- Starai VJ, Escalante-Semerena JC (2004) Acetyl-coenzyme A synthetase (AMP forming). Cell Mol Life Sci 61: 2020-2030.
- Fairchild TA, Patejunas G (1999) Cloning and expression profile of human inorganic pyrophosphatase. Biochim Biophys Acta 1447: 133-136.
- Yang Y, Cai J, Yin J, Wang D, Bai Z, Zhang J, et al. (2015) Inorganic pyrophosphatase (*PPA1*) is a negative prognostic marker for human gastric cancer. Int J Clin Exp Pathol 8: 12482-12490.
- Zambrowicz BP, Abuin A, Ramirez-Solis R, Richter LJ, Piggott J, et al. (2003) Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. Proc Natl Acad Sci USA 100: 14109-14114.
- Panda H, Pandey RS, Debata PR, Supakar PC (2007) Cloning, expression, and functional analysis of rat liver cytosolic inorganic pyrophosphatase gene and characterization of its functional promoter. Gene Expr 14: 13-22.
- Panda H, Pandey RS, Debata PR, Supakar PC (2007) Age-dependent differential expression and activity of rat liver cytosolic inorganic pyrophosphatase gene. Biogerontology 8: 517-525.
- Thierry-Mieg D, Thierry-Mieg J (2006) AceView: A comprehensive cDNA-supported gene and transcripts annotation. Genome Biology 7: S12.
- Pohjanjoki P, Lahti R, Goldman A, Cooperman BS (1998) Evolutionary conservation of enzymatic catalysis: quantitative comparison of the effects of mutation of aligned residues in *Saccharomyces cerevisiae* and *Escherichia coli* inorganic pyrophosphatases on enzymatic activity. Biochemistry 37: 1754-1761.
- Altschul F, Vyas V, Cornfield A, Goodin S, Ravikumar TS, et al. (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, et al. (2003) The human genome browser at UCSC. Genome Res 12: 994-1006.
- Sievers F, Higgins DG (2014) Clustal omega. Curr Protoc Bioinformatics 48: 1-16.
- Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl Acad Sci USA 103: 1412-1417.
- Kopp J, Schwede T (2004) The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. Nucleic Acids Res 32: D230-D234.
- Heikinheimo P, Lehtonen J, Baykov A, Lahti R, Cooperman BS, et al. (1996) The structural basis for pyrophosphatase catalysis. Structure 4: 1491-508.
- Marchler-Bauer A, Panchenko AR, Ariel N, Bryant SH (2002) Comparison of sequence and structure alignments for protein domains. Proteins 48: 439-446.
- GTEX Consortium (2015) Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348: 648-660.
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465-469.
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.
- Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol 537: 113-37.
- Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, et al. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol 21: 566-569.
- Oksanen E, Ahonen AK, Tuominen H, Tuominen V, Lahti R, et al. (2007) A complete structural description of the catalytic cycle of yeast pyrophosphatase. Biochemistry 46: 1228-1239.
- Colak G, Xie Z, Zhu AY, Dai L, Lu Z, et al. (2013) Identification of lysine succinylation substrates and the succinylation regulatory enzyme CobB in *Escherichia coli*. Mol Cell Proteomics 12: 3509-3520.
- Simon MC (1993) Transcription factor GATA-1 and erythroid development. Proc Soc Exp Biol Med 202: 115-121.
- Hoang T, Lambert JA, Martin R (2016) SCL/TAL1 in Hematopoiesis and Cellular Reprogramming. Curr Top Dev Biol 118: 163-204.
- Putt W, Palmén J, Nicaud V, Tregouet DA, Tahri-Daizadeh N, et al. (2004) Variation in USF1 shows haplotype effects, gene : gene and gene : environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. Hum Mol Genet 13: 1587-1597.
- Ma M, Wu W, Li Q, Li J, Sheng Z, Shi J, et al. (2015) N-myc is a key switch regulating the proliferation cycle of postnatal cerebellar granule cell progenitors. Sci Rep 5: 12740.
- Krock BL, Eisinger-Mathason TS, Giannoukos DN, Shay JE, Gohil M, et al. (2015) The aryl hydrocarbon receptor nuclear translocator is an essential regulator of murine hematopoietic stem cell viability. Blood 125: 3263-3272.
- Wu S, Liu W, Zhou L (2016) MiR-590-3p regulates osteogenic differentiation of human mesenchymal stem cells by regulating APC gene. Biochem Biophys Res Commun 478: 1582-1587.
- Polewski MD, Johnson KA, Foster M, Millán JL, Terkeltaub R (2010) Inorganic pyrophosphatase induces type I collagen in osteoblasts. Bone 46: 81-90.
- Yi YJ, Sutovsky M, Kennedy C, Sutovsky P (2012) Identification of the inorganic pyrophosphate metabolizing, ATP substituting pathway in mammalian spermatozoa. PLoS One 7: e34524.
- Opazo JC, Butts GT, Nery MF, Storz JF, Hoffmann FG (2013) Whole genome duplication and the functional diversification of teleost fish hemoglobins. Mol Biol Evol 30: 140-153.