CBB 520 HW #1

Guangjian Du

1) What is the size of the human IPTA protein (in both kilo Daltons and amino-acid length)?

**Answer:**

21.4 kDa; 194 amino-acid length.

2) What is the main result reported in this paper (three sentence maximum)?

**Answer:**

Genetic variants leading to inosine triphosphatase deficiency, a condition not thought to be clinically important, protect against haemolytic anaemia in hepatitis-C-infected patients receiving RBV.

Two known functional variants conferring reduced ITPA activity are responsible for the protection against anaemia identified in the original GWAS.

Drug responses are also likely to be heavily influenced by variants that are too rare for effective representation in GWAS.

3) What does ITPA stand for?

**Answer:**

Inosine triphosphate pyrophosphatase (also called: ITPase).

4) What is the enzyme commission number for ITPA? What are the substrates and product of this enzyme?

**Answer:**

EC: 3.6.1.19.

Substrates: inosine triphosphate and deoxyinosine triphosphate.

Products: monophosphate nucleotide and diphosphate.



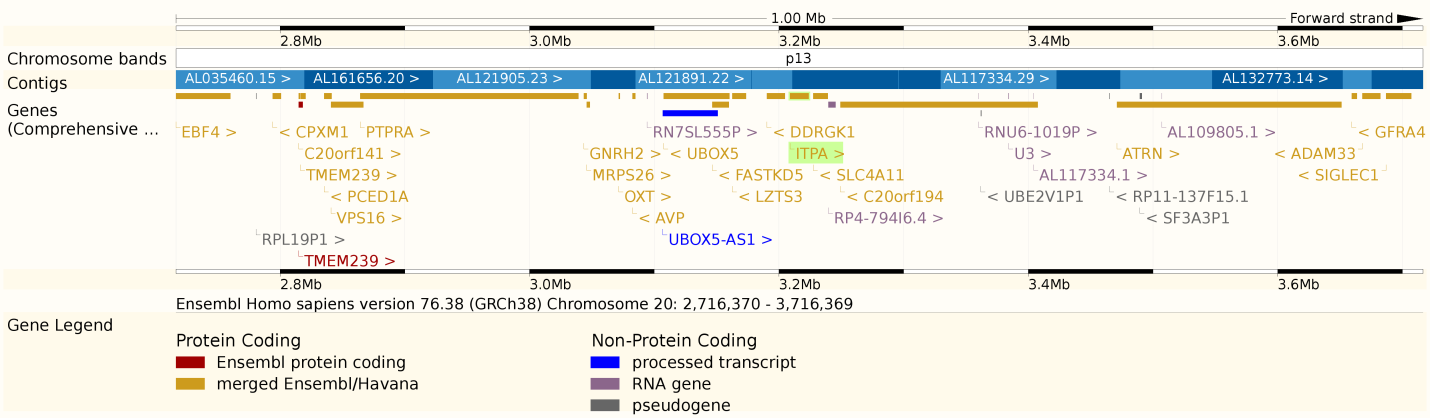
5) Where in the genome is it encoded? (i.e. chromosome and location, map locus). Also, how far from the centromere is this gene (in kb)?

**Answer:**

Chromosome 20: 3208868-3223870.

Distance from centromere: 23212 kb.

Map locus:



5) What is the OMIM number for this human gene?

**Answer:**

OMIM: 147520.

6) Are there any human diseases known to result from mutations in this gene, if so what? (2 sentence max)

**Answer:**

Diseases: cancer, psychiatric, pulmonary langerhans's cell histiocytosis, human fibroblast.

7) Describe GWAS (3 sentence maximum)

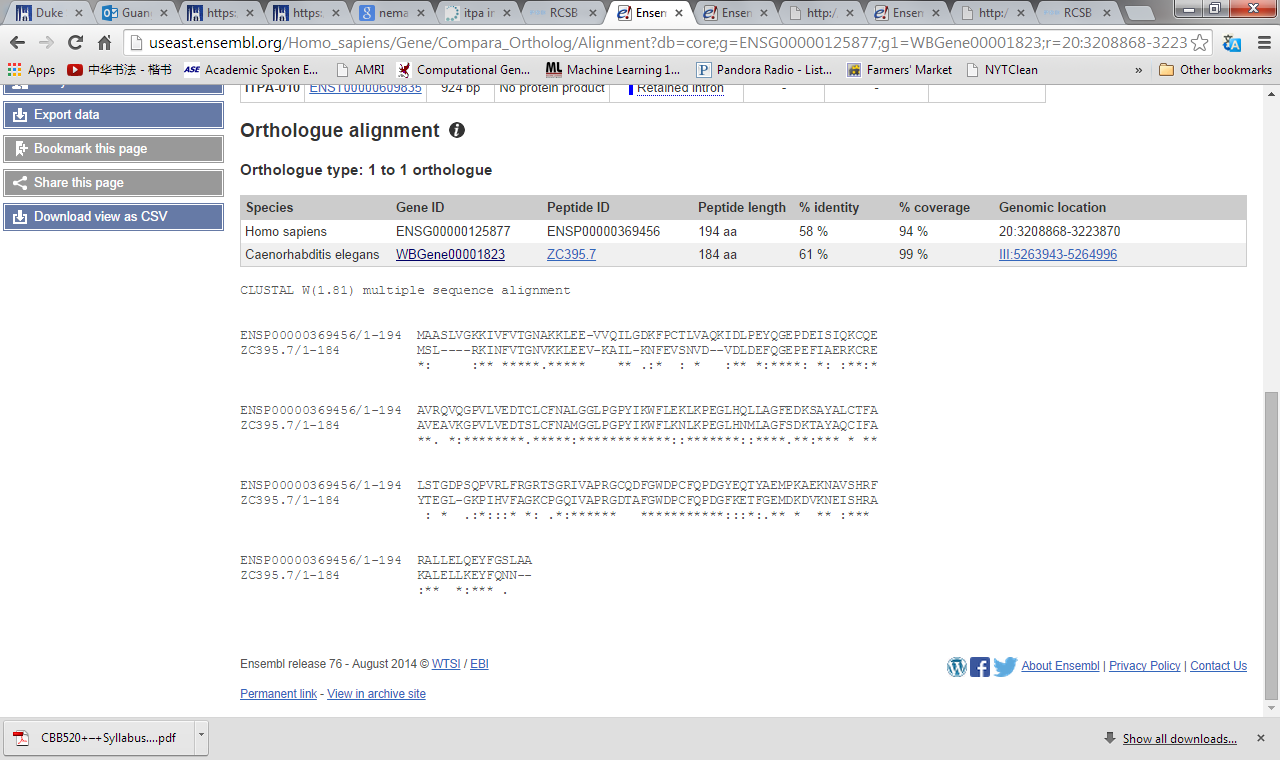
**Answer:**

GWAS: Genome-wide Association Study; It is an examination of many common genetic variants in different individuals to see if any variant is associated with a trait. GWAS typically focus on associations between single-nucleotide polymorphisms (SNPs) and traits like major diseases.

8) How is the human ITPA protein and nematode (C. elegans) homolog? (Percent identity)

**Answer:**

Homolog: 61% identity.



9) How similar are the lengths of the human and nematode genes?

**Answer:**

Lengths similar for ITPA gene:

Human ITPA gene length: *3,208,868-3,223,870* = 15002 bp.

Nematode ITPA gene length: *5,263,943-5,264,996* = 1053 bp.

The lengths are quite different.

Lengths of the whole genome:

Human: about 3 billion bp.

Nematode: about 97 million bp.

10) What is the refseq number for this protein?

**Answer:**

Protein refseq: NP\_258412

*3 alternative transcripts:* [*NP\_001254552.1*](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=protein&cmd=Search&doptcmdl=GenPept&term=NP_001254552.1)[*NP\_258412.1*](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=protein&cmd=Search&doptcmdl=GenPept&term=NP_258412.1)[*NP\_852470.1*](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=protein&cmd=Search&doptcmdl=GenPept&term=NP_852470.1)

*cDNA clone refseq: NM\_033453*

11) How many introns are in this human gene?

**Answer:**

Introns: 16.

<http://www.ncbi.nlm.nih.gov/gene/3704>

12) List 3 SNPs from dbSNP found within exons of the human ITPA gene and three SNPs found within introns.

**Answer:**

Within Introns:

[rs7270101](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7270101), rs8114386, rs6107258

Within Exons:

[rs1127354](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1127354), [rs6139034](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=6139034) , rs9101

[*http://www.ncbi.nlm.nih.gov/snp/?term=ITPA&SITE=NcbiHome&submit=Go*](http://www.ncbi.nlm.nih.gov/snp/?term=ITPA&SITE=NcbiHome&submit=Go)[*http://www.ncbi.nlm.nih.gov/projects/SNP/*](http://www.ncbi.nlm.nih.gov/projects/SNP/)

13) How many CpG islands lie within 5,000 base pairs upstream and downstream of this gene?

**Answer:**

2770

[*http://www.bioinformatics.org/sms2/cpg\_islands.html*](http://www.bioinformatics.org/sms2/cpg_islands.html)

14) Are there splicing variants for human ITPA gene? Multiple polyadenylation sites?

**Answer:**

Yes, there are 10 splicing variants: ITPA\_001--ITPA\_010.

Yes, there are several Multiple polyadenylation sites.

http://useast.ensembl.org/Homo\_sapiens/Gene/Splice?db=core;g=ENSG00000125877;r=20:3208868-3223870

15) List three tissue/cell types that ITPA genes are highly expressed in human.

**Answer:**

Tissues: heart, liver, pancreas;

16) What are other proteins that ITPA have been shown to interact with?

**Answer:**

1: [UBC](http://thebiogrid.org/113164/summary/homo-sapiens/ubc.html) | HMG20 polyubiquitin-C

2: [WNT4](http://thebiogrid.org/119939/summary/homo-sapiens/wnt4.html) | RP1-224A6.7, WNT-4, SERKAL

3: [VPS4B](http://thebiogrid.org/114901/summary/homo-sapiens/vps4b.html) | MIG1, SKD1, SKD1B, VPS4-2 vacuolar protein sorting 4 homolog B (S. cerevisiae)

4: [SERPINB11](http://thebiogrid.org/124598/summary/homo-sapiens/serpinb11.html) | SERPIN11, EPIPIN serpin peptidase inhibitor, clade B (ovalbumin)

5: [TSHR](http://thebiogrid.org/113104/summary/homo-sapiens/tshr.html) | LGR3, CHNG1, hTSHR-I thyroid stimulating hormone receptor

6: [ENO1](http://thebiogrid.org/108338/summary/homo-sapiens/eno1.html) | PPH, ENO1L1, NNE, MPB1 2-phospho-D-glycerate hydro-lyase

7: [GLO1](http://thebiogrid.org/109001/summary/homo-sapiens/glo1.html) | GLYI, GLOD,1S-D-lactoylglutathione methylglyoxal lyase

<http://thebiogrid.org/109909>

16) List 3 (by accession number) sequences containing parts of the coding region of the ITPA gene of the domestic cat from the Trace archive database Approximately how many coding region sequences are there in the trace archive from the domestic cat for this gene?

**Answer:**

Accession number:

Ti: 2217773510

Ti: 1266288079

Ti: 251818822

<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=TraceArchive&PAGE_TYPE=BlastSearch&PROG_DEFAULTS=on>

<http://useast.ensembl.org/Felis_catus/Location/View?g=ENSFCAG00000013414;r=A3:29095224-29105664;t=ENSFCAT00000013418>

17) What is the Swiss-prot number for the human ITPA gene?

**Answer:**

UniProtKB/Swiss-Prot: [ITPA\_HUMAN, Q9BY32](http://www.uniprot.org/uniprot/Q9BY32)

18) What is the PIR number for the human ITPA protein?

**Answer:**

PIR#: PR: 000009153; PR: Q9BY32

19) On what mouse chromosome is the mouse ITPA gene?

**Answer:**

Mouse ITPA gene: Chromosome 2: 130,667,610-130,681,614

20) Who first cloned a human cDNA for ITPA? (give reference)

**Answer:**

Shengrong Lin, etc

Reference: Cloning, expression, and characterization of a human inosine triphosphate pyrophosphatase encoded by the itpa gene, *J Biol Chem*. 2001 Jun 1;276(22):18695-701.

21) What are the Molecular Function, Biological Process, and Cellular Component Go Annotations for this gene; and the evidence code for each?

**Answer:**

|  |  |
| --- | --- |
| Biological Process | code |
| ITP catabolic process | IEA |
| nucleotide metabolic process | IEA |
| nucleoside triphosphate catabolic process | IEA |
| deoxyribonucleoside triphosphate catabolic process | IEA |
| small molecule metabolic process | TAS |
| chromosome organization | IEA |
| nucleobase-containing small molecule metabolic process | TAS |
| Molecular Function | code |
| nucleotide binding | IEA |
| hydrolase activity | IEA |
| dITP diphosphatase activity | IEA & EXP |
| ITP diphosphatase activity | EXP |
| XTP diphosphatase activity | EXP |
| metal ion binding | IEA |
| nucleoside-triphosphate diphosphatase activity | IEA |
| GO Cellular Component | code |
| cytoplasm | IEA |
| cytosol | TAS |

<http://www.uniprot.org/uniprot/Q9BY32>  
<http://www.ebi.ac.uk/QuickGO/GProtein?ac=Q9BY32>

22) List one of the STS markers found in the human ITPA gene. What is an STS marker

(2 sentence maximum)

**Answer:**

STS marker: STS-AA026396.

STS marker: The Sequence-Tagged Site (STS) is a relatively short, easily PCR-amplified sequence (200 to 500 bp) which can be specifically amplified by [PCR](http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechPCR.shtml) and detected in the presence of all other genomic sequences and whose location in the genome is mapped.

23) What is the unigene number for the human ITPA gene? What is a unigene? (2 sentence maximum)

**Answer:**

UGID: 222716.     UniGene Hs.415299

UniGene is an [NCBI](http://en.wikipedia.org/wiki/National_Center_for_Biotechnology_Information) database of the [transcriptome](http://en.wikipedia.org/wiki/Transcriptome) and thus, despite the name, not primarily a database for[genes](http://en.wikipedia.org/wiki/Genes). Each entry is a set of [transcripts](http://en.wikipedia.org/wiki/Transcription_(genetics)) that appear to stem from the same [transcription](http://en.wikipedia.org/wiki/Transcription_(genetics)) locus (i.e. gene or expressed [pseudogene](http://en.wikipedia.org/wiki/Pseudogene)). Information on [protein](http://en.wikipedia.org/wiki/Protein) similarities, gene expression, [cDNA](http://en.wikipedia.org/wiki/Complementary_DNA) [clones](http://en.wikipedia.org/wiki/Molecular_cloning), and genomic location is included with each entry.

24) What is the clinical phenotype they were investigating in this paper?

**Answer:**

The threshold of Hb reduction of >3g dl-1 was chosen as a clinically significant Hb decline.

25) What is the name of the Saccharomyces cerevisiae (yeast) homolog of ITPA? How was a mutation in this gene first found, and what is the deletion phenotype (3 sentence maximum)

**Answer:**

Name: HAM1.

When HAM1 was mutated the yeast became supersensitive to the mutagenic effects of the purine analog 6-N-hydroxylaminopurine (HAP).

Deletion phenotype: HAM1-1.

26) Where did the authors estimate was their success rate of genotyping calls? (Percentage)

**Answer:**

At step 1: Infinium BeadStudio Raw Data Analysis.

When re-clustering step creates SNP calling errors, the authors identified a procedure to prevent the errant calls from being released in the final report.

The success rate of genotyping calls is 97.5%.