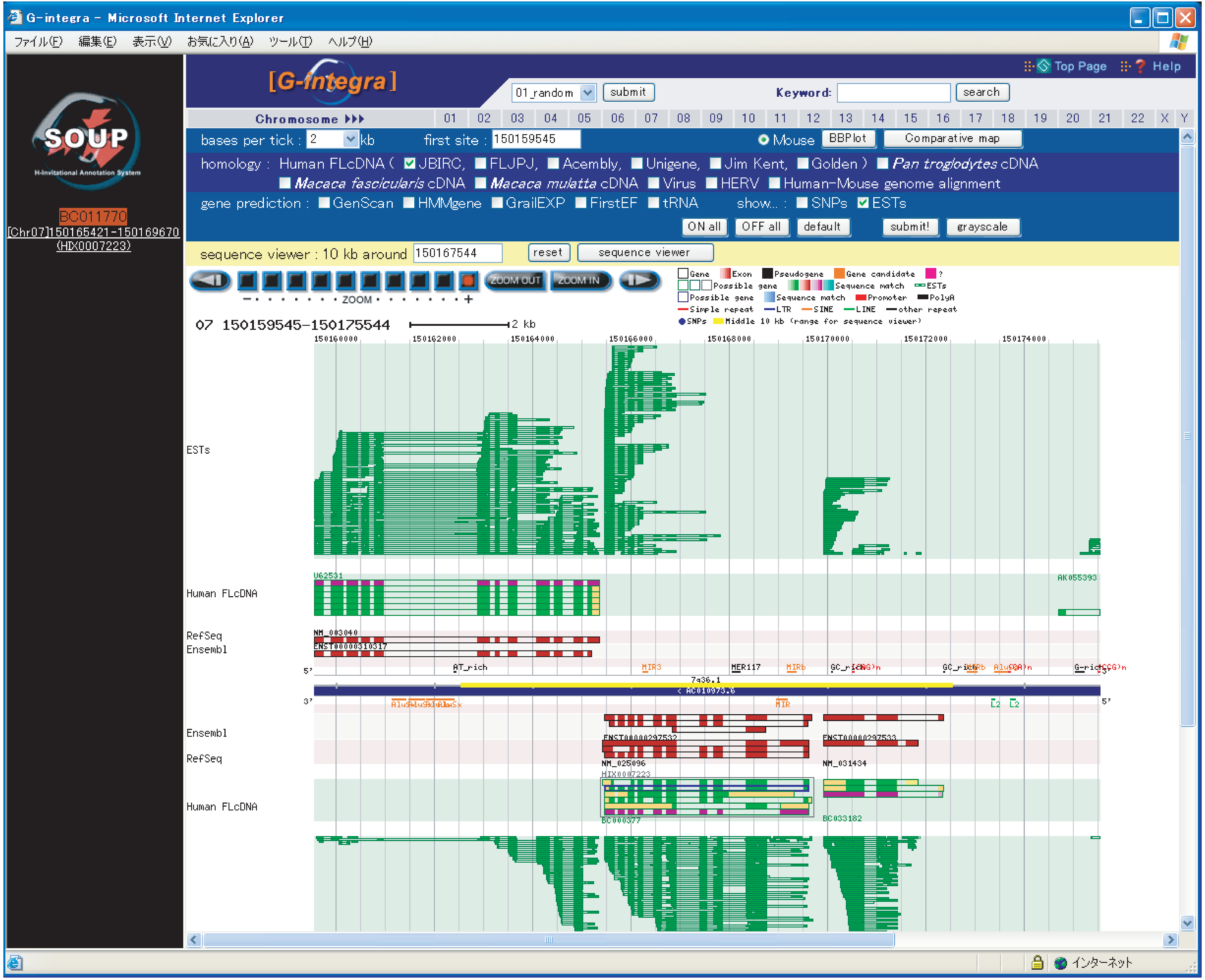
Figure S9. H-Inv Annotation Viewers

(A) G-integra: A gemone mapping viewer.



G-integra - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

BC011770
[Chr07]150165421-150169670
(HIX0007223)

01_random submit Keyword: search

bases per tick : 2 kb first site : 150159545 ● Mouse

homology : Human FLcDNA (☑ JBIRC, ☐ FLJPJ, ☐ Acembly, ☐ Unigene, ☐ Jim Kent, ☐ Golden) ☐ *Pan troglodytes* cDNA
☐ *Macaca fascicularis* cDNA ☐ *Macaca mulatta* cDNA ☐ Virus ☐ HERV ☐ Human-Mouse genome alignment

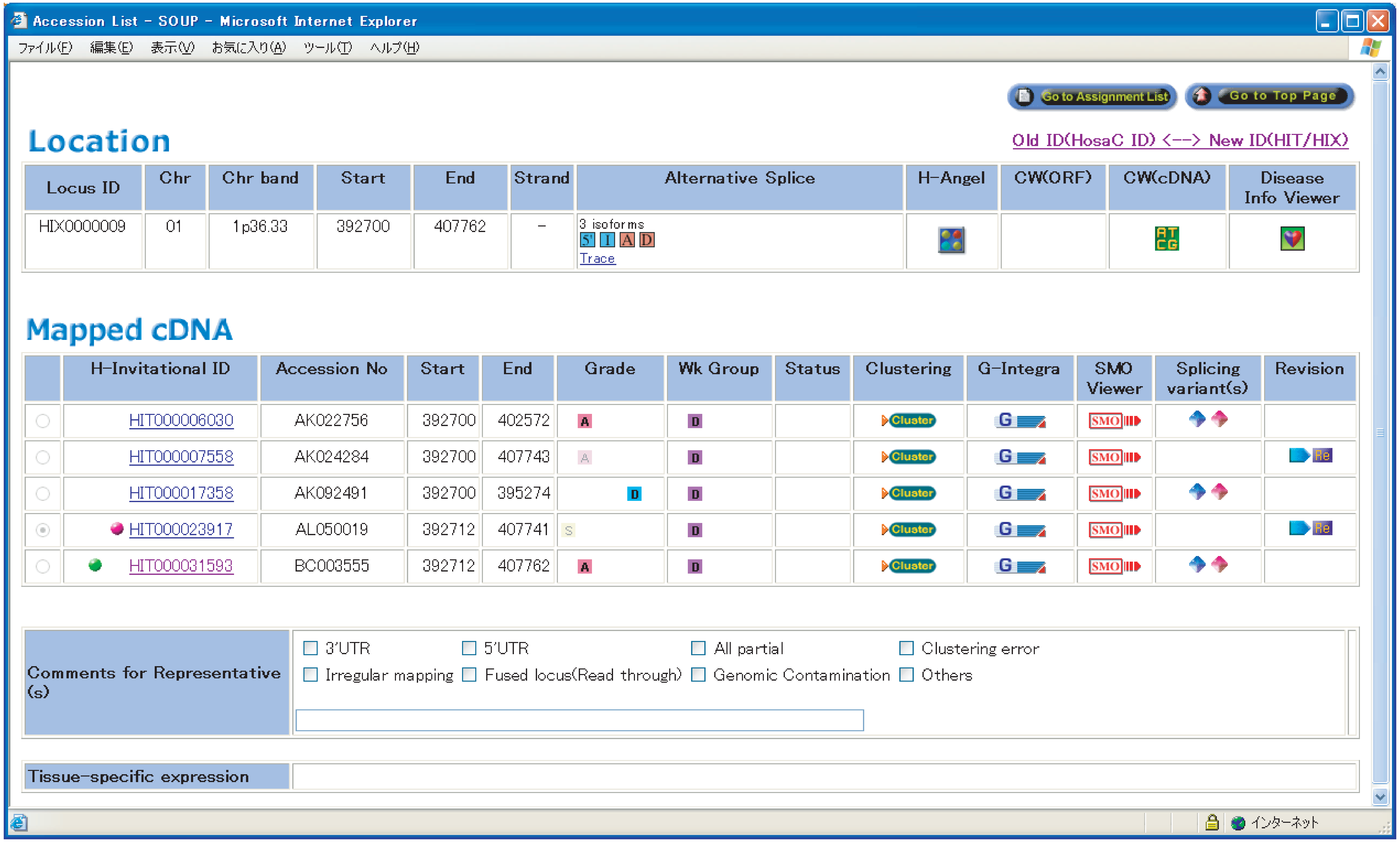
gene prediction : ☐ GenScan ☐ HMMgene ☐ GrailEXP ☐ FirstEF ☐ tRNA show... : ☐ SNPs ☑ ESTs

sequence viewer : 10 kb around 150167544

07 150159545-150175544

Figure S9. H-Inv Annotation Viewers

(B) SOUP Locus annotation Viewer.



Location

Old ID(HosaC ID) <--> New ID(HIT/HIX)

Locus ID	Chr	Chr band	Start	End	Strand	Alternative Splice	H-Angel	CW(ORF)	CW(cDNA)	Disease Info Viewer
HIX0000009	01	1p36.33	392700	407762	-	3 isoforms 5' I A D Trace			ATCG	

Mapped cDNA

	H-Invitational ID	Accession No	Start	End	Grade	Wk Group	Status	Clustering	G-Integra	SMO Viewer	Splicing variant(s)	Revision
○	HIT000006030	AK022756	392700	402572	A	D		Cluster	G	SMO		
○	HIT000007558	AK024284	392700	407743	A	D		Cluster	G	SMO		Re
○	HIT000017358	AK092491	392700	395274	D	D		Cluster	G	SMO		
●	● HIT000023917	AL050019	392712	407741	S	D		Cluster	G	SMO		Re
○	● HIT000031593	BC003555	392712	407762	A	D		Cluster	G	SMO		

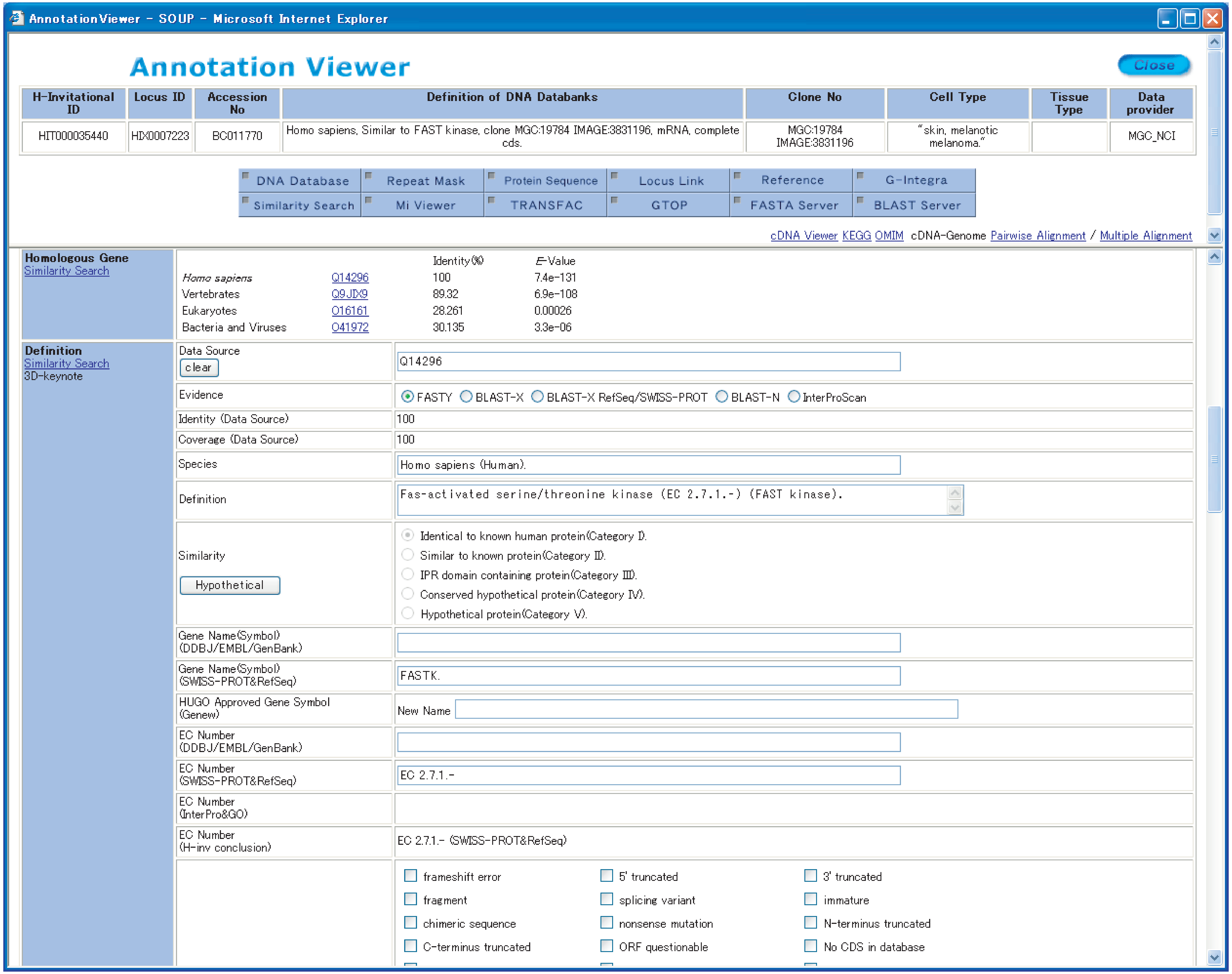
Comments for Representative (s)

☐ 3'UTR ☐ 5'UTR ☐ All partial ☐ Clustering error
☐ Irregular mapping ☐ Fused locus(Read through) ☐ Genomic Contamination ☐ Others

Tissue-specific expression

Figure S9. H-Inv Annotation Viewers

(C) SOUP cDNA annotation Viewer.



H-Invitational ID	Locus ID	Accession No	Definition of DNA Databanks	Clone No	Cell Type	Tissue Type	Data provider
HIT000035440	HIX0007223	BC011770	Homo sapiens, Similar to FAST kinase, clone MGC:19784 IMAGE:3831196, mRNA, complete cds.	MGC:19784 IMAGE:3831196	"skin, melanotic melanoma."		MGC_NCI

Homologous Gene
Similarity Search

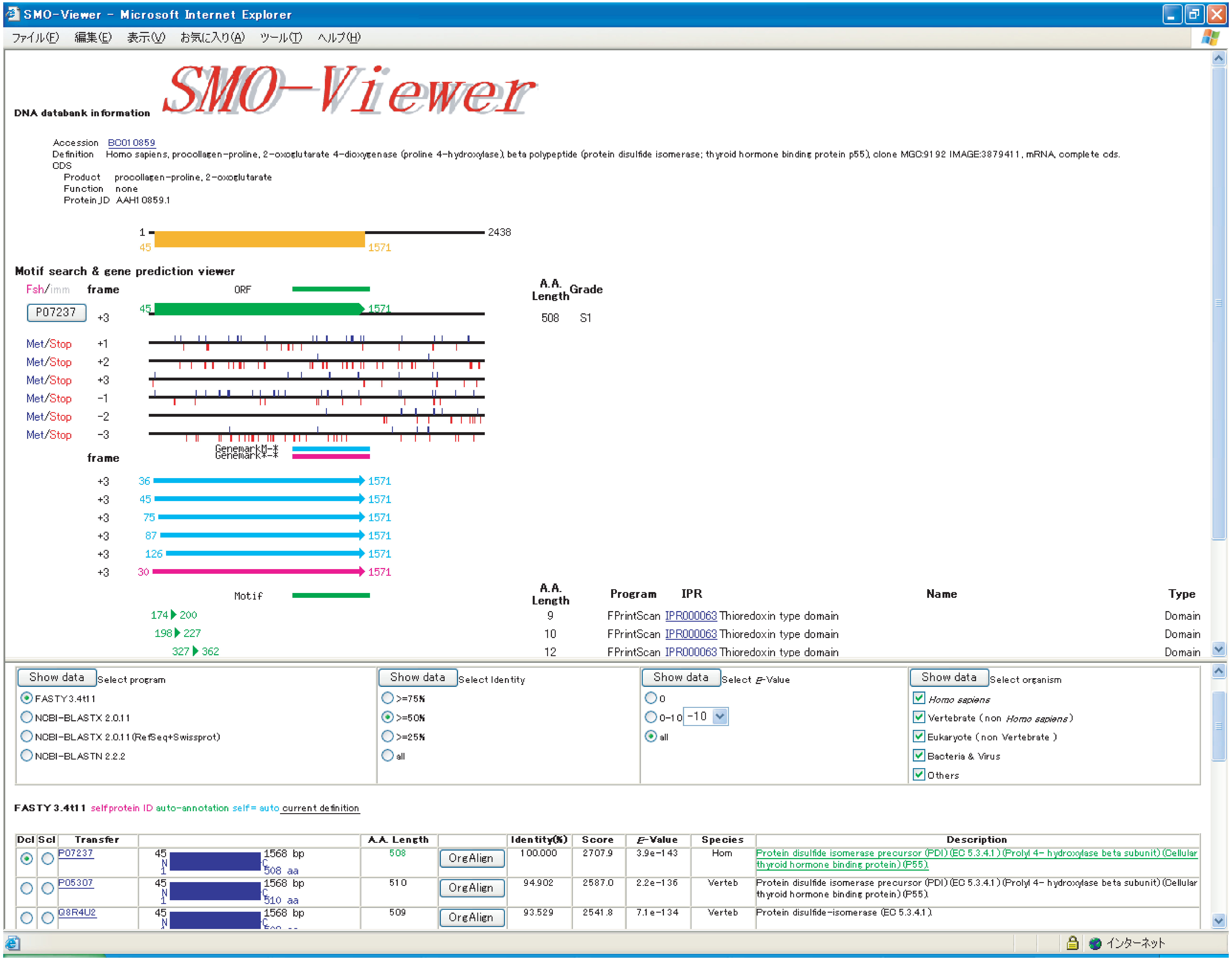
		Identity(%)	*E*-Value
Homo sapiens	Q14296	100	7.4e-131
Vertebrates	Q9JIX9	89.32	6.9e-108
Eukaryotes	O16161	28.261	0.00026
Bacteria and Viruses	O41972	30.135	3.3e-06

Definition
Similarity Search
3D-keynote

Data Source clear	Q14296
Evidence	● FASTY ○ BLAST-X ○ BLAST-X RefSeq/SWISS-PROT ○ BLAST-N ○ InterProScan
Identity (Data Source)	100
Coverage (Data Source)	100
Species	Homo sapiens (Human).
Definition	Fas-activated serine/threonine kinase (EC 2.7.1.-) (FAST kinase).
Similarity Hypothetical	● Identical to known human protein(Category I). ○ Similar to known protein(Category II). ○ IPR domain containing protein(Category III). ○ Conserved hypothetical protein(Category IV). ○ Hypothetical protein(Category V).
Gene Name(Symbol) (DDBJ/EMBL/GenBank)	
Gene Name(Symbol) (SWISS-PROT&RefSeq)	FASTK.
HUGO Approved Gene Symbol (Genew)	New Name
EC Number (DDBJ/EMBL/GenBank)	
EC Number (SWISS-PROT&RefSeq)	EC 2.7.1.-
EC Number (InterPro&GO)	
EC Number (H-inv conclusion)	EC 2.7.1.- (SWISS-PROT&RefSeq)
	☐ frameshift error ☐ 5' truncated ☐ 3' truncated ☐ fragment ☐ splicing variant ☐ immature ☐ chimeric sequence ☐ nonsense mutation ☐ N-terminus truncated ☐ C-terminus truncated ☐ ORF questionable ☐ No CDS in database

Figure S9. H-Inv Annotation Viewers

(D) SMOViewer: The similarity, motif and ORF information viewer.



DNA databank information

Accession BC010859
Definition Homo sapiens, procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55), clone MGC:9192 IMAGE:3879411, mRNA, complete cds.
CDS
Product procollagen-proline, 2-oxoglutarate
Function none
Protein_ID AAH10859.1

A.A. Length	Grade
508	S1

frame	start	end
+3	36	1571
+3	45	1571
+3	75	1571
+3	87	1571
+3	126	1571
+3	30	1571

Motif	A.A. Length	Program	IPR	Name	Type
174 ▶ 200	9	FPrintScan	IPR000063	Thioredoxin type domain	Domain
198 ▶ 227	10	FPrintScan	IPR000063	Thioredoxin type domain	Domain
327 ▶ 362	12	FPrintScan	IPR000063	Thioredoxin type domain	Domain

Select program
● FASTY3.4t11
○ NCBI-BLASTX 2.0.11
○ NCBI-BLASTX 2.0.11 (RefSeq+Swissprot)
○ NCBI-BLASTN 2.2.2

Select Identity
○ >=75%
● >=50%
○ >=25%
○ all

Select E-Value
○ 0
○ 0-10 -10
● all

Select organism
☑ *Homo sapiens*
☑ Vertebrate (non *Homo sapiens*)
☑ Eukaryote (non Vertebrate)
☑ Bacteria & Virus
☑ Others

FASTY 3.4t11 self protein ID auto-annotation self= auto current definition

Dcl	Scl	Transfer		A.A. Length		Identity(%)	Score	E-Value	Species	Description
●	○	P07237	45 N 1 – 1568 bp C 508 aa	508	OrgAlign	100.000	2707.9	3.9e-143	Hom	Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase beta subunit) (Cellular thyroid hormone binding protein) (P55).
○	○	P05307	45 N 1 – 1568 bp C 510 aa	510	OrgAlign	94.902	2587.0	2.2e-136	Verteb	Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase beta subunit) (Cellular thyroid hormone binding protein) (P55).
○	○	Q8R4U2	45 N – 1568 bp C	509	OrgAlign	93.529	2541.8	7.1e-134	Verteb	Protein disulfide-isomerase (EC 5.3.4.1).