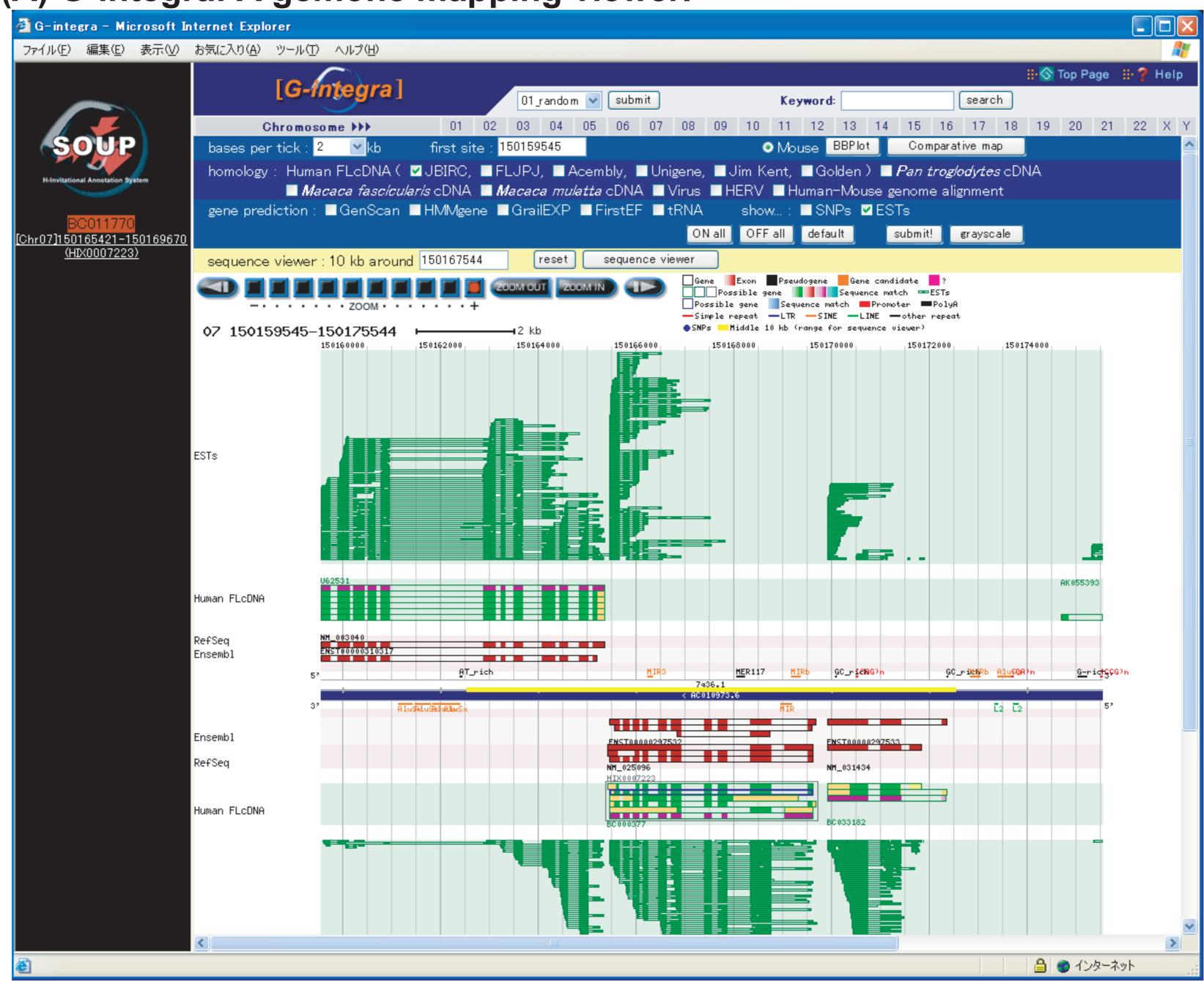
# **Figure S9. H-Inv Annotation Viewers**

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



# Figure S9. H-Inv Annotation Viewers

## (B) SOUP Locus annotation Viewer.

🕘 Access	ion List	- SOUP	– Microsoft	Internet	Explorer
ファイル(F)	編集(E)	表示(V)	お気(: 入り(A)	- ツール(T)	ヘルプ(H)

ion											Old ID(	HosaC ID)	<> New	ID(HIT/HIX)
Chr	Chr I	band	Start	End	Strar	nd	Alternative S	Splice	H-A	ngel	CW(OR	F) CW	(cDNA)	Disease Info Viewer
9 01	1 p36	3.33	392700	407762	-	3 isoforms <b>5 1 A D</b> <u>Trace</u>			2				AT CG	V
		Acce	ssion No	Start	End	Grade	Wk Group	Status	Clustering	G-J	Integra	SMO Viewer		
<u>HIT000006</u>	030	AK	022756	392700	402572	A	D		Cluster		G	SMO	• •	
HIT000007	558	AK	024284	392700	407743	A	D		Cluster		G	SMO		
HIT000017	358	AK	092491	392700	395274	D	D		Cluster		G	SMO	• •	
HIT000023	<u>917</u>	AL	050019	392712	407741	S	D		Cluster		G	SMO		Re Re
HIT000031	<u>593</u>	BC	003555	392712	407762	A	D		Cluster		G	SMO	• •	
											error			
tor Repres	sentativ	ve	] Irregular m	apping 🔲	Fused loci	us(Read throu	gh) 📃 Genomi	c Contamin	nation 📃 Oth	ers				
	01         d cDN         witational         HIT000007         HIT000017         HIT000031	01       1 p36         d cDNA         witational ID         HIT000006030         HIT000017358         HIT000031593	01       1p36.33         d cDNA         witational ID       Access         HIT000006030       AK         HIT000017358       AK         HIT000023917       AL         HIT000031593       BC	01       1 p36.33       392700         d cDNA       392700         Accession No         HT000006030       AK022756         HT00007558       AK024284         HT000017358       AK024291         HT0000317       AL050019         HT000031593       BC03555	3       01       1p36.33       392700       407762         d cDNA         No         No	01       1 p36.33       392700       407762       -         d cDNA         Notitational ID       Accession No       Start       End         HIT000006030       AK022756       392700       402572         HIT000007558       AK024284       392700       407743         HIT000017358       AK092491       392700       395274         HIT000023917       AL050019       392712       407741         HIT000031593       BC003555       392712       407762	3       01       1 p36.33       392700       407762       -       3 isoforms in a bit a	a       01       1p36.33       392700       407762       -       3 isoforms Incce         C CONA         Accession No       Start       End       Grade       Wk Group         HIT00000630       AK022756       392700       407743       A       In         HIT000017558       AK024284       392700       407743       A       In         HIT000023917       AL050019       392712       407741       In       In         HIT000031593       BC003555       392712       407762       A       In	a       01       1 p36.33       392700       407762       -       3 isoforms Image         a       image       image	B       O1       1 p36.33       392700       407762       -       3 isoforms I I I I I I I I I I I I I I I I I I I	B       O1       1 p36.33       392700       407762       -       3 isoforms I acc	B       O1       1 p36.33       392700       407762       -       3 isoforms inclusion       Image: Constant inclusion       Image: Consta	0       01       1 p36.33       392700       407762       -       3 isoforms recent and rec	n         1

<mark>Acce</mark> :  ァイル( <u>F</u> )					ernet Explore ール(T) ヘルプ(											
													<b>6</b> 60 to	Assignment L	ist) 🚯 🐠	o to Top Page
Lo	catio	n											Old ID(	HosaC ID	) <> Nev	VID(HIT/HIX)
Lo	cus ID	Chr	Chr	band	Start	End	Strar	ıd	Alternative	Splice		H-Ang	el CW(OR	F) CW	(cDNA)	Disease Info Viewer
HIXO	)000009	01	1 p36	8.33	392700	407762	-	3 isoforms <b>5 1 A D</b> <u>Trace</u>				<b>22</b>			AT CG	<b>V</b>
Ма	pped	<b>CDN</b> tational		Acco	ssion No	Start	End	Grade	Wk Group	Status	Cluet	tering	G-Integra	SMO	Splicing	Revision
	11 11101	ta tionai		ACCE		otart	LIIU	Grade		otatus	Olusi	tering	G Integra	Viewer	variant(s	
0	HI	T000006	030	AK	022756	392700	402572	A	D		<b>PCI</b>	uster	G	SMO	**	
0	HI	T000007	<u>'558</u>	AK	024284	392700	407743	A	D		► CI	uater	G	SMO		
0	HI	T000017	' <u>358</u>	AK	092491	392700	395274	D	D		<b>▶C</b> I	uster	G	SMO	- 🔶 🔶	
0	) <u>H</u> I	T000023	1 <u>917</u>	AL	050019	392712	407741	S	D		►CI	uəter	G	SMO		
0	<u>е</u> н	T000031	<u>593</u>	BC	003555	392712	407762	A	D		►CI	uster	G	SMO	**	
Comments for Representative (s)       3'UTR       5'UTR       All partial       Clustering error																
issu	e-specifi	c expre	ession													
																🧿 インターネット

Comments for Representative (s)	<ul> <li>3'UTR</li> <li>Irregular mapping</li> </ul>	5'UTR Fused locus(Read through)	<ul> <li>All partial</li> <li>Genomic C</li> </ul>
Tissue-specific expression			

# Figure S9. H-Inv Annotation Viewers

### (C) SOUP cDNA annotation Viewer.

	Anno	otatio	n Viev	Ne	er					
H-Invitational ID	Locus ID	Accession No			Definition	of DNA Databanks	S			
HIT000035440	HDX0007223	BC011770	Homo sapiens, 3	Simila	ar to FAST kinase,	clone MGC:19784 IMA cds.	GE:383	1196, m		
		E DN	A Database	•	Repeat Mask	Protein Sequence	•	Locu		
		■ Simi	larity Search	F	Mi Viewer	TRANSFAC	=	G		
lomologous Gen	e				Identity (%)	) <i>E</i> -Value				
<u>Similarity Search</u>		mo sapiens	Q1429	6	100	7.4e-131				
		rtebrates	<u>Q9JIX</u>	_	89.32	6.9e-108				
	Eul	karyotes	01616	_	28.261	0.00026				
	Ba	cteria and Viruse	es <u>04197</u>	2	30.135	3.3e-06				
Definition	Dat	a Source			014206					
<u>iimilarity Search</u> D-keynote	Cl	ear			Q14296					
,	Evi	dence			● FASTY ● BLAST-X ● BLAST-X RefSeq/SWIS					
	Ider	ntity (Data Sourc	e)		100					
	Cov	erage (Data Sou	irce)		100					
	Spe	cies			Homosapiens (	(Human).				
	Def	inition			Fas-activated serine/threonine kinase (					
	Sim	ilarity			<ul> <li>Identical to known human protein(Category D).</li> <li>Similar to known protein(Category D).</li> </ul>					
	_	•				containing protein(Cate		۲D.		
		Hypothetical	J			ypothetical protein(Ca				
						protein(Category V).				
		e Name(Symbol) BJ/EMBL/GenB								
		e Name(Symbol) /ISS-PROT&RefS			FASTK.					
		GO Approved Ge new)	ne Symbol		New Name					
		Number )BJ/EMBL/GenB	Bank)							
	(SW	Number /ISS-PROT&RefS	Geq)		EC 2.7.1					
	ant	Number erPro&GO)								
		Number inv conclusion)			EC 2.7.1 (SWISS	6-PROT&RefSeq)				
					🔲 frameshift e	rror	_	' truncat		
					🔲 fragment		📃 sp	plicing		
					📃 chimeric sed	quence	📃 no	onsense		
					C-terminus	4		RF que		

								X
								^
							Close	
		Clone No		Cell Typ	e	Tissue Type	Data provider	
mRNA, comple	te	MGC:19784 IMAGE:3831196	i	″skin, melar melanoma	notic ».‴		MGC_NCI	≡
us Link		Reference	•	G-Integra				
ТОР	•	FASTA Server	BL	AST Server				
		<u>cDNA Viewer KE</u>	<u> 66 OI</u>	<u>MIM</u> cDNA-Genor	ne <u>Pairwise</u>	Alignment / M	ultiple Alignment	>
								^
SS-PROT 🔘	BL	AST-N 🔘 InterProSc	an					
(EC 2.7.1.	-)	(FAST kinase).		~				
ated		🔲 3' trur	ncated					
variant		📃 immat						
e mutation		N-ter	minus t	runcated				
estionable			DS in d	atabase				~

## Figure S9. H-Inv Annotation Viewers (D) SMOViewer: The similarity, motif and ORF information viewer.

ファイル(E) 編集(E) 表示(M) お気に入り(A) ツール(E) ヘルプ(E) SMO-Viewer	
SMO-Viewer	
Accession <u>BC010859</u> 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 cd: CDS Product procollagen-proline, 2-oxoglutarate Function none Protein JD AAH1 0859.1	s.
1 - 2438 45 1571	
Motif search & gene prediction viewer	
Fsh/imm     frame     ORF       P07237     45	
+3 +3 508 S1	
Met/Stop +2	
Met/Stop -1	
Met/Stop -2	
Met/Stop -3	
frame Genemark#-#	
+3 36 1571	
+3 45 1571	
+3 75 1571	
+3 87 1571 +3 126 1571	
+3 30 1571	
Motif A.A. Program IPR Name	Туре
174 200 9 FPrintScan <u>IPR000063</u> Thioredoxin type domain	Domain
198 227 10 FPrintScan <u>IPR000063</u> Thioredoxin type domain	Domain
327 362 12 FPrintScan IPR000063 Thioredoxin type domain	Domain
Show data Select program Show data Select Identity Show data Select E-Value Show data Select organism	
● FASTY3.4t11         ● >=75%         ● 0	
○ NCBI-BLASTX 2.0.11     ○ >=50%     ○ 0-10     -10     ✓	
○ NCBI-BLASTX 2.0.11 (RefSeq+Swissprot)       ○ >=25%         ○ NCBI-BLASTX 2.0.11 (RefSeq+Swissprot)       ○ >=25%	
O NCBI-BLASTN 2.2.2 Image: Sector a & Virus	
FASTY 3.4t11 selfprotein ID auto-annotation self= auto current definition	
Del Sel Transfer A.A. Length Identity(%) Score F-Value Species Description	
OrgAlign     100.000     2707.9     3.9e-143     Hom     Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase bet     N	ta subunit) (Cellular
Image: Portage in the system of the syste	
O       P07237       45       1568 bp       508       OrgAlign       1 00.000       2707.9       3.9e-1.43       Hom       Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase bet thyroid hormone binding protein) (P55).         O       P05307       45       508 aa       510       OrgAlign       94.902       2587.0       2.2e-136       Verteb       Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase bet thyroid hormone binding protein) (P55).	

🚰 SMO-Viewer – Microsoft Internet Explorer							∎₽⊻
ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)							
DNA databank in formation Accession <u>BC010859</u> Definition Homo sapiens, procollagen-proline, 2-oxoglutarate 4-dioxyge CDS Product procollagen-proline, 2-oxoglutarate Function none			de (protein di	isulfide isomers	ise; thyroid hor	mone binding protein p55), clone MGC:9192 IMAGE:3879411, mRNA, complete cds.	
Motif search & gene prediction viewer Fsh/imm frame ORF	24 571 571	A.A. Length					
P07237       +3         Met/Stop       +1         Met/Stop       +2         Met/Stop       +3         Met/Stop       -1         Met/Stop       -2         Met/Stop       -3         frame       Genemark#-*		508	51				
+3 45 11 +3 75 11 +3 87 11 +3 126 11	571 571 571 571 571 571	A.A. Length	Pro	ogram II	PR	Name	Туре
174 200		9		intScan <u>IPRO</u>	00063 Thiored	loxin type domain	Domain
198 227		10				loxin type domain	Domain
327 362 Show data Select program • FASTY 3.4t11 • NOBI-BLASTX 2.0.11 • NOBI-BLASTX 2.0.11 (RefSeq+Swissprot) • NOBI-BLASTN 2.2.2	Show data >=75% >=50% >=25% all	entity	F Pr		lata <sub>Select</sub>	loxin type domain E-Value Nomo sapiens Vertebrate (non Homo sapiens) Eukaryote (non Vertebrate) Bacteria & Virus Others	Domain
FASTY 3.4t11 selfprotein ID auto-annotation self = auto current definition         Dcl Sci       Transfer	A Length	Identity(%)	Score	<i>E</i> -Value	Species	Description	
O         P07237         45         1568 bp           N         1         508 aa           O         P05307         45           N         1568 bp	508 OrgAlign 510 OrgAlign	94.902	2707.9	3.9e-143 2.2e-136	Hom Verteb	Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase beta subu thyroid hormone binding protein) (P55). Protein disulfide isomerase precursor (PDI) (EC 5.3.4.1) (Prolyl 4- hydroxylase beta subu thyroid hormone binding protein) (P55).	
1 510 aa 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	509 OrgAlign	93.529	2541.8	7.1e-134		Protein disulfide-isomerase (EC 5.3.4.1 ).	
	Constituent						<b>Հ</b> ահ
							ተሆ <u>ኮ</u>