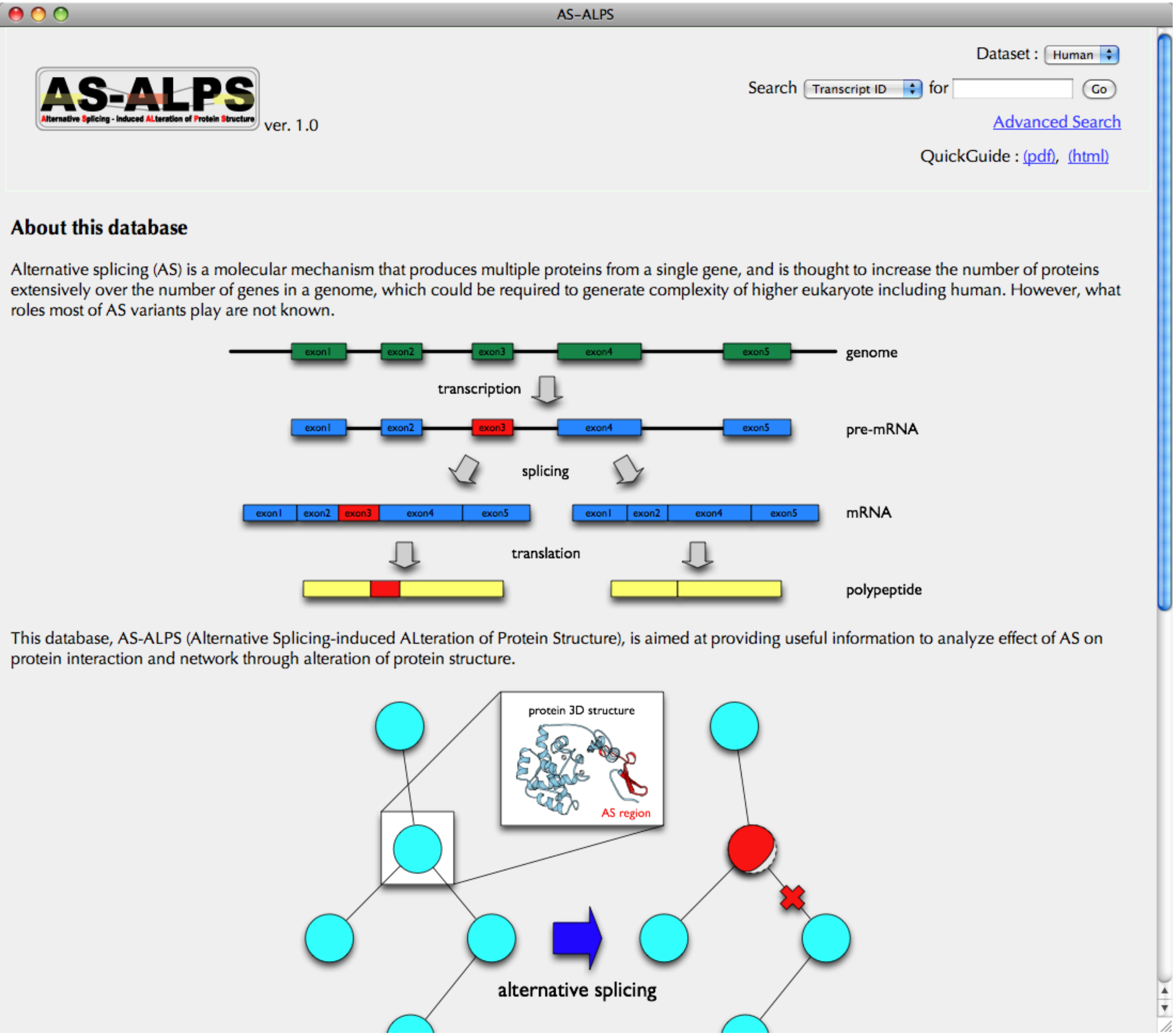
Quick Guide of AS-ALPS

Front Page



Dataset : Human

Search Transcript ID for

QuickGuide : (pdf), (html)

About this database

Alternative splicing (AS) is a molecular mechanism that produces multiple proteins from a single gene, and is thought to increase the number of proteins extensively over the number of genes in a genome, which could be required to generate complexity of higher eukaryote including human. However, what roles most of AS variants play are not known.

This database, AS-ALPS (Alternative Splicing-induced ALteration of Protein Structure), is aimed at providing useful information to analyze effect of AS on protein interaction and network through alteration of protein structure.

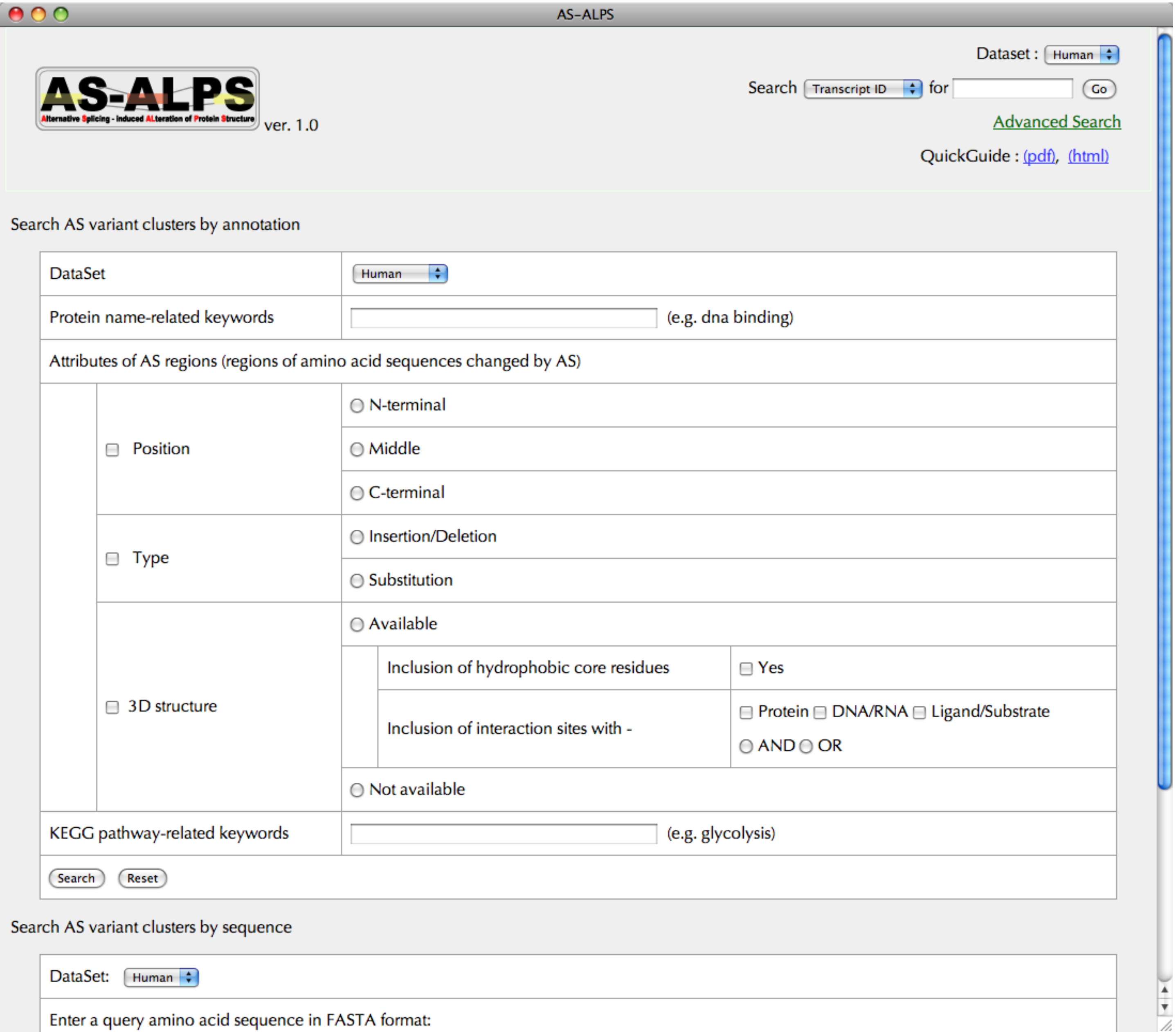
On the search form in the top right corner, you can search for a cluster of transcripts produced by AS from a single gene, in either human or mouse, specified by either Entrez Gene ID (e.g., 10681) or Transcript ID in H-InvDB, FANTOM or RefSeq (e.g., NM_016194).

→ Go to Page 3

Alternatively, you can move to "Advanced Search" page by clicking the hyper-link below the search box.

→ Go to Page 2

Advanced Search Page 1



Dataset : Human

Search Transcript ID for

QuickGuide : (pdf), (html)

Search AS variant clusters by annotation

DataSet		Human	
Protein name-related keywords		(e.g. dna binding)	
Attributes of AS regions (regions of amino acid sequences changed by AS)			
☐ Position		○ N-terminal	
		○ Middle	
		○ C-terminal	
☐ Type		○ Insertion/Deletion	
		○ Substitution	
☐ 3D structure		○ Available	
		Inclusion of hydrophobic core residues	☐ Yes
		Inclusion of interaction sites with -	☐ Protein ☐ DNA/RNA ☐ Ligand/Substrate ○ AND ○ OR
		○ Not available	
KEGG pathway-related keywords		(e.g. glycolysis)	

Search AS variant clusters by sequence

DataSet: Human

Enter a query amino acid sequence in FASTA format:

1. Search by annotation

You can combine the search conditions below.

(1) human or mouse

(2) protein name

(3) attributes of AS regions:
- (a) position
- (b) type
- (c) if 3D info is available,
  - (i) whether AS regions include hydrophobic core residues
  - (ii) what kind of interaction sites AS regions include

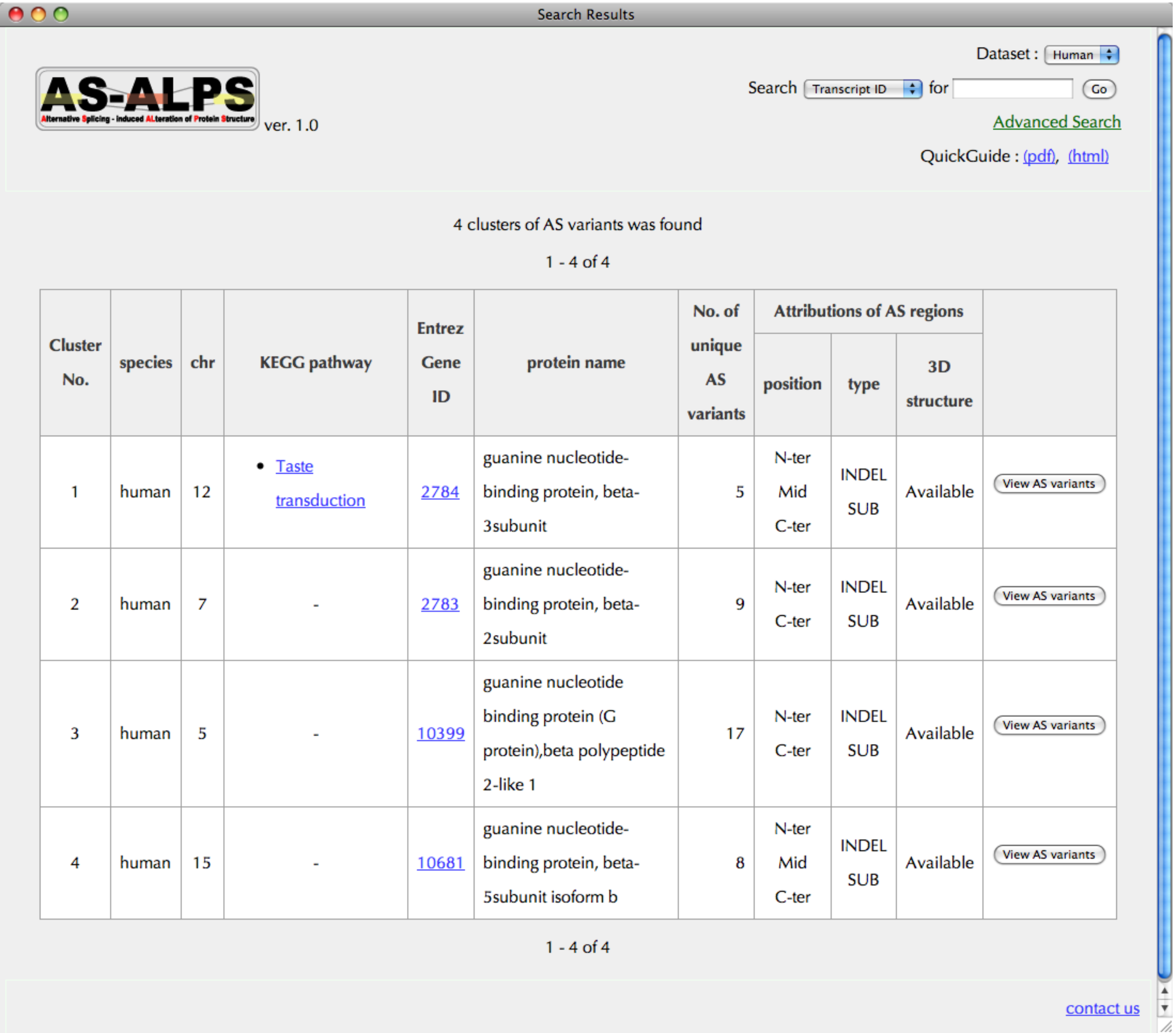
(4) KEGG pathway name

→ Go to Page 3

2. Search by sequence

→ Go to Page 12

Search Results Page



Dataset : Human

Search Transcript ID for

QuickGuide : (pdf), (html)

4 clusters of AS variants was found

1 - 4 of 4

Cluster No.	species	chr	KEGG pathway	Entrez Gene ID	protein name	No. of unique AS variants	Attributions of AS regions			
							position	type	3D structure	
1	human	12	- Taste transduction	2784	guanine nucleotide-binding protein, beta-3subunit	5	N-ter Mid C-ter	INDEL SUB	Available	View AS variants
2	human	7	-	2783	guanine nucleotide-binding protein, beta-2subunit	9	N-ter C-ter	INDEL SUB	Available	View AS variants
3	human	5	-	10399	guanine nucleotide binding protein (G protein),beta polypeptide 2-like 1	17	N-ter C-ter	INDEL SUB	Available	View AS variants
4	human	15	-	10681	guanine nucleotide-binding protein, beta-5subunit isoform b	8	N-ter Mid C-ter	INDEL SUB	Available	View AS variants

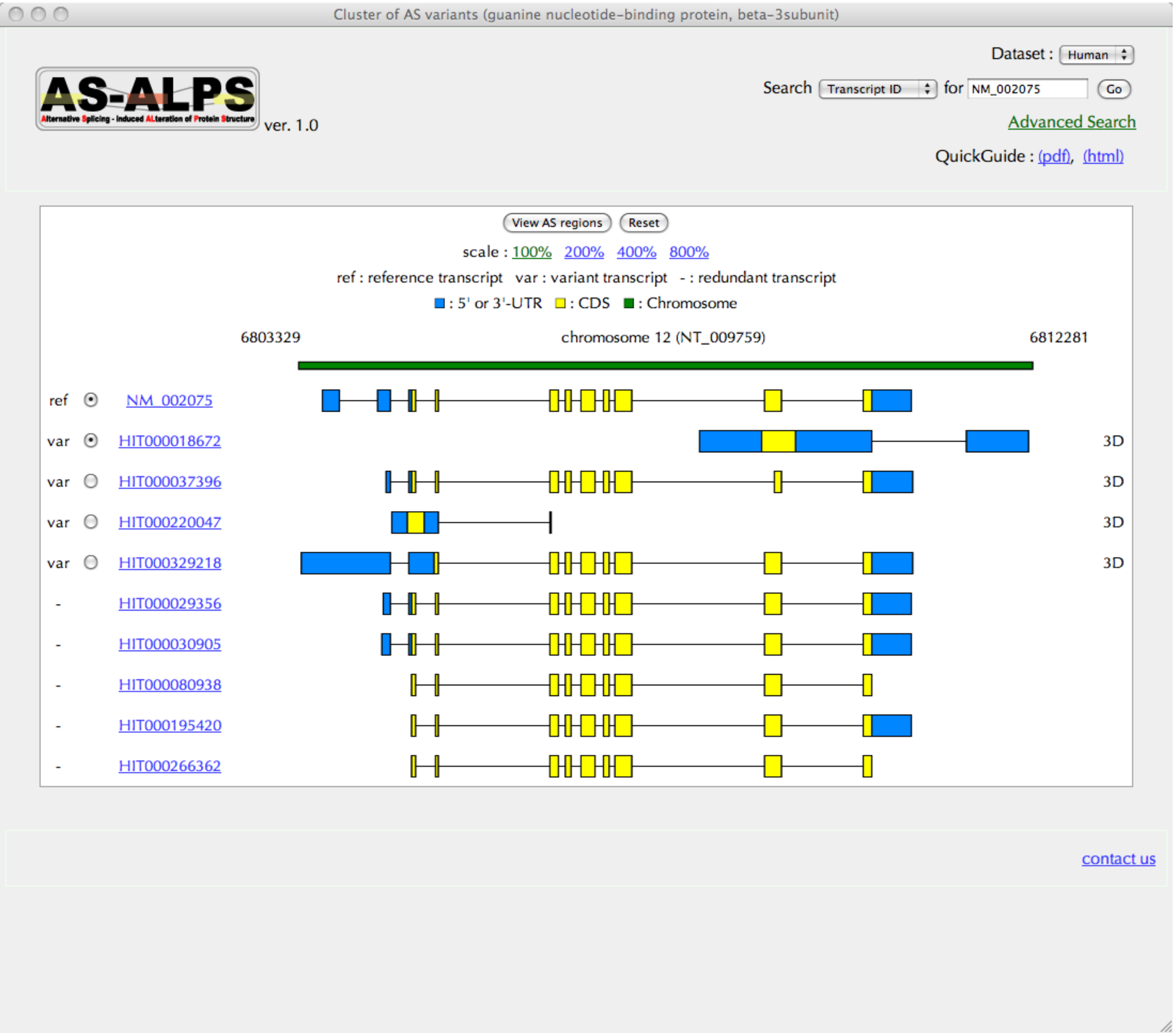
1 - 4 of 4

In this table of search results, annotated info is briefly shown for each cluster of transcripts produced by AS from a single gene.

Click a button in the rightmost column to view AS variants in a cluster.

→ Go to Page 4

AS variant cluster Page



Cluster of AS variants (guanine nucleotide-binding protein, beta-3subunit)

ver. 1.0

Dataset : Human

Search Transcript ID for NM_002075 Go

scale : 100% 200% 400% 800%

ref : reference transcript var : variant transcript - : redundant transcript

🟦 : 5' or 3'-UTR 🟨 : CDS 🟩 : Chromosome

6803329 chromosome 12 (NT_009759) 6812281

ref	●	NM_002075	
var	●	HIT000018672	3D
var	○	HIT000037396	3D
var	○	HIT000220047	3D
var	○	HIT000329218	3D
-		HIT000029356	
-		HIT000030905	
-		HIT000080938	
-		HIT000195420	
-		HIT000266362	

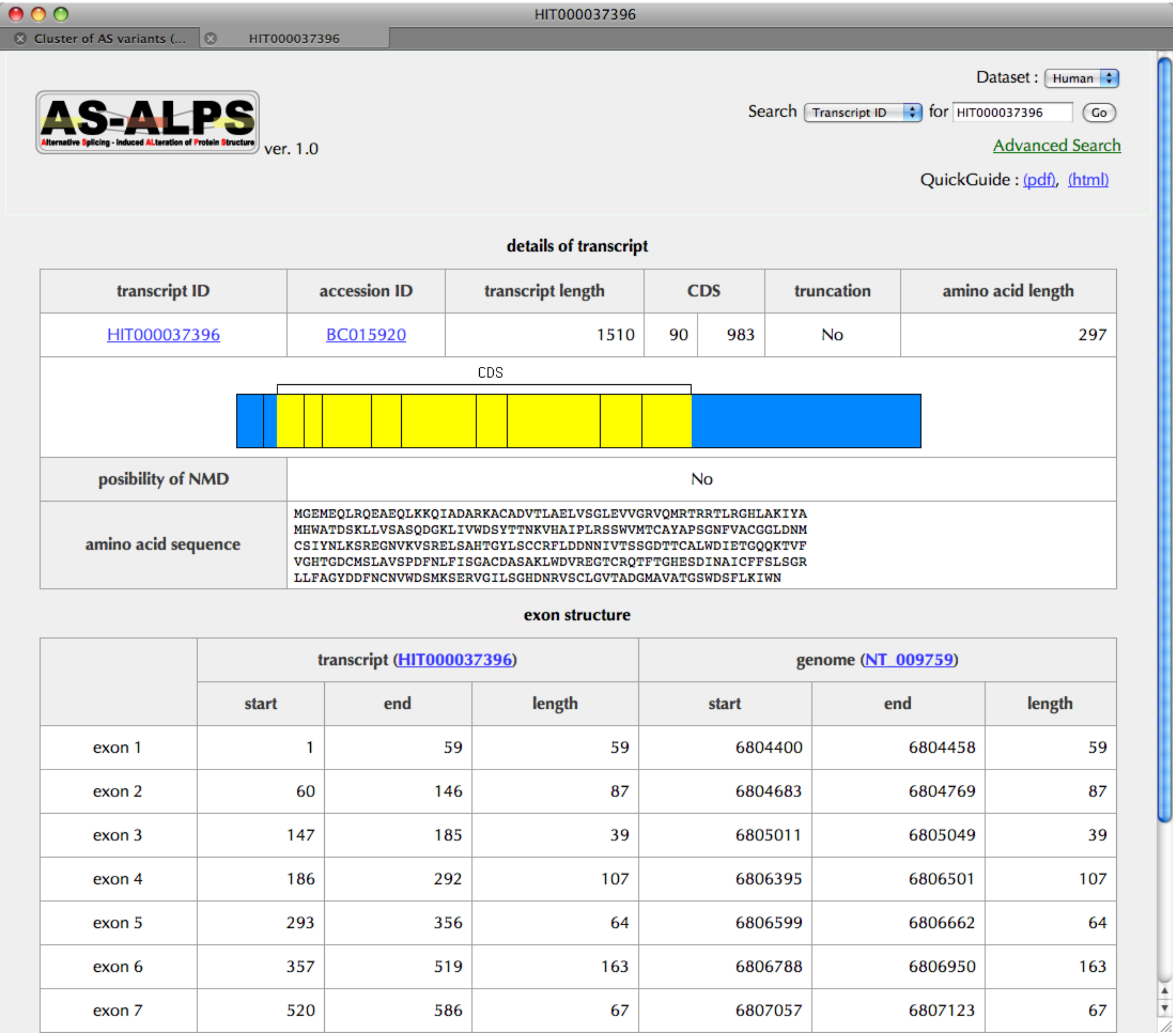
You can see here a schematic representation of alignments between a chromosome and transcripts.

AS regions are detected by a comparison between a pre-defined reference transcript and one of other variant transcripts.

Select one variant transcript in the leftmost column and push the "View AS regions" button.
→ Go to Page 6

Click a transcript ID to see its details
→ Go to Page 5

Transcript Details Page



details of transcript

transcript ID	accession ID	transcript length	CDS		truncation	amino acid length
HIT000037396	BC015920	1510	90	983	No	297

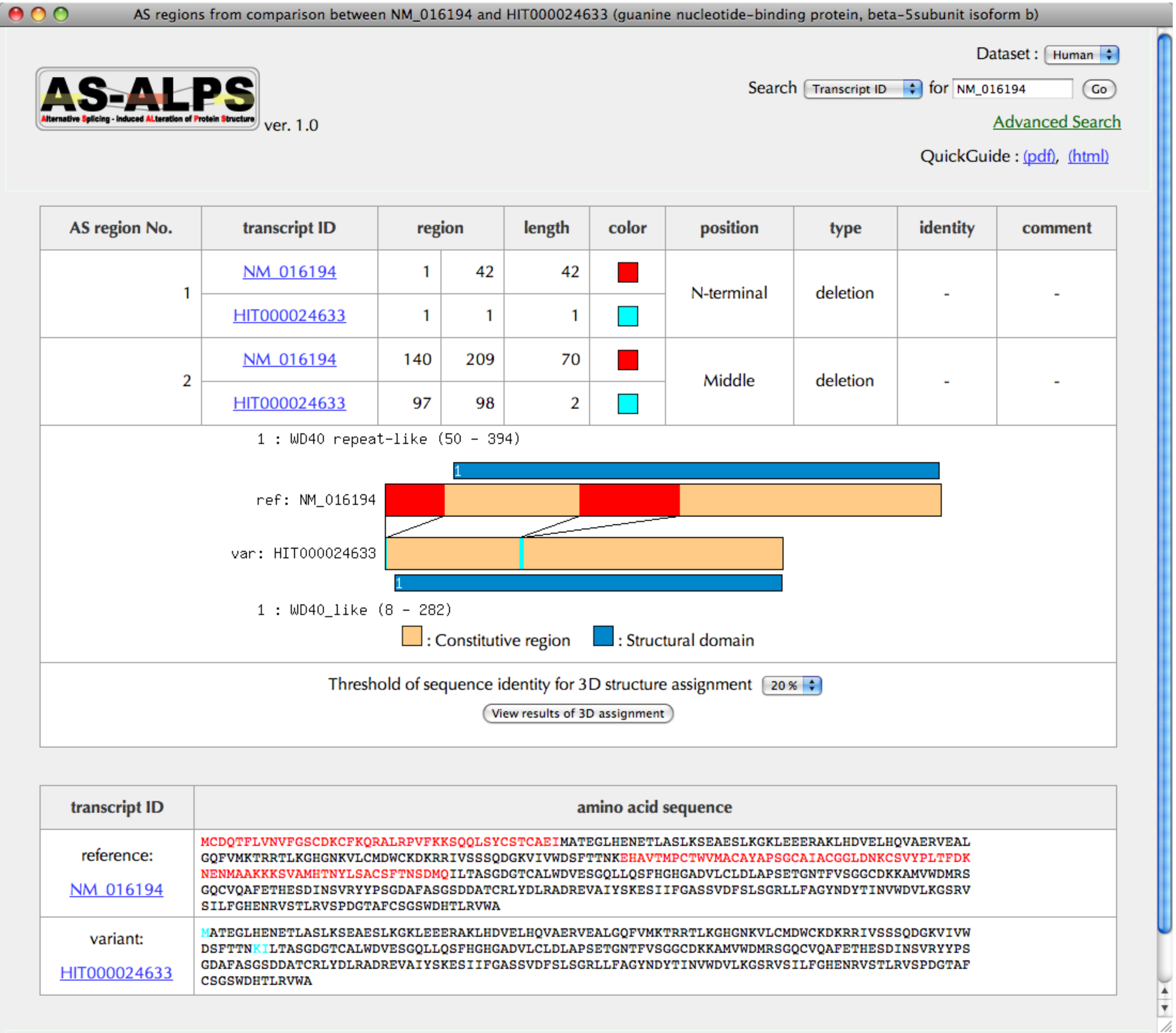
posibility of NMD	No
amino acid sequence	MGEMEQLRQEAEQLKKQIADARKACADVTLAELVSGLEVVGRVQMRTRRTLRGHLAKIYA MHWATDSKLLVSASQDGKLIVWDSYTTNKVHAIPLRSSWVMTCAYAPSGNFVACGGLDNM CSIYNLKSREGNVKVSRELSAHTGYLSCCRFLDDNNIVTSSGDTTCALWDIETGQQKTVF VGHTGDCMSLAVSPDFNLFISGACDASAKLWDVREGTCRQTFTGHESDINAICFFSLSGR LLFAGYDDFNCNVWDSMKSERVGILSGHDNRVSCLGVTADGMAVATGSWDSFLKIWN

exon structure

	transcript (HIT000037396)			genome (NT_009759)		
	start	end	length	start	end	length
exon 1	1	59	59	6804400	6804458	59
exon 2	60	146	87	6804683	6804769	87
exon 3	147	185	39	6805011	6805049	39
exon 4	186	292	107	6806395	6806501	107
exon 5	293	356	64	6806599	6806662	64
exon 6	357	519	163	6806788	6806950	163
exon 7	520	586	67	6807057	6807123	67

For each transcript requested, you can check the region of CDS (coding sequence), the status of truncation at the 5’ and 3’ termini, a possibility of NMD (nonsense-mediated mRNA decay) judged from the sequence, the translated amino acid sequence and the exon composition.

AS regions View Page



AS regions from comparison between NM_016194 and HIT000024633 (guanine nucleotide-binding protein, beta-5subunit isoform b)

Dataset : Human

Search Transcript ID for NM_016194 Go

AS region No.	transcript ID	region		length	color	position	type	identity	comment
1	NM_016194	1	42	42	🟥	N-terminal	deletion	-	-
	HIT000024633	1	1	1	🟦				
2	NM_016194	140	209	70	🟥	Middle	deletion	-	-
	HIT000024633	97	98	2	🟦				

🟧 : Constitutive region 🟦 : Structural domain

Threshold of sequence identity for 3D structure assignment 20 %

transcript ID	amino acid sequence
reference: NM_016194	MCDQTFLVNVFGSCDKCFKQRALRPVFKKSQQLSYCSTCAEIMATEGLHENETLASLKSEAESLKGKLEEERAKLHDVELHQVAERVEAL GQFVMKTRRTLKGHGNKVLCMDWCKDKRRIVSSSQDGKVIVWDSFTTNKEHAVTMPCTWVMACAYAPSGCAIACGGLDNKCSVYPLTFDK NENMAAKKKSVAMHTNYLSACSFTNSDMQILTASGDGTCALWDVESGQLLQSFHGHGADVLCLDLAPSETGNTFVSGGCDKKAMVWDMRS GQCVQAFETHESDINSVRYYPSGDAFASGSDDATCRLYDLRADREVAIYSKESIIFGASSVDFSLSGRLLFAGYNDYTINVWDVLKGSRV SILFGHENRVSTLRVSPDGTAFCSGSWDHTLRVWA
variant: HIT000024633	MATEGLHENETLASLKSEAESLKGKLEEERAKLHDVELHQVAERVEALGQFVMKTRRTLKGHGNKVLCMDWCKDKRRIVSSSQDGKVIVW DSFTTNKILTASGDGTCALWDVESGQLLQSFHGHGADVLCLDLAPSETGNTFVSGGCDKKAMVWDMRSGQCVQAFETHESDINSVRYYPS GDAFASGSDDATCRLYDLRADREVAIYSKESIIFGASSVDFSLSGRLLFAGYNDYTINVWDVLKGSRVSILFGHENRVSTLRVSPDGTAF CSGSWDHTLRVWA

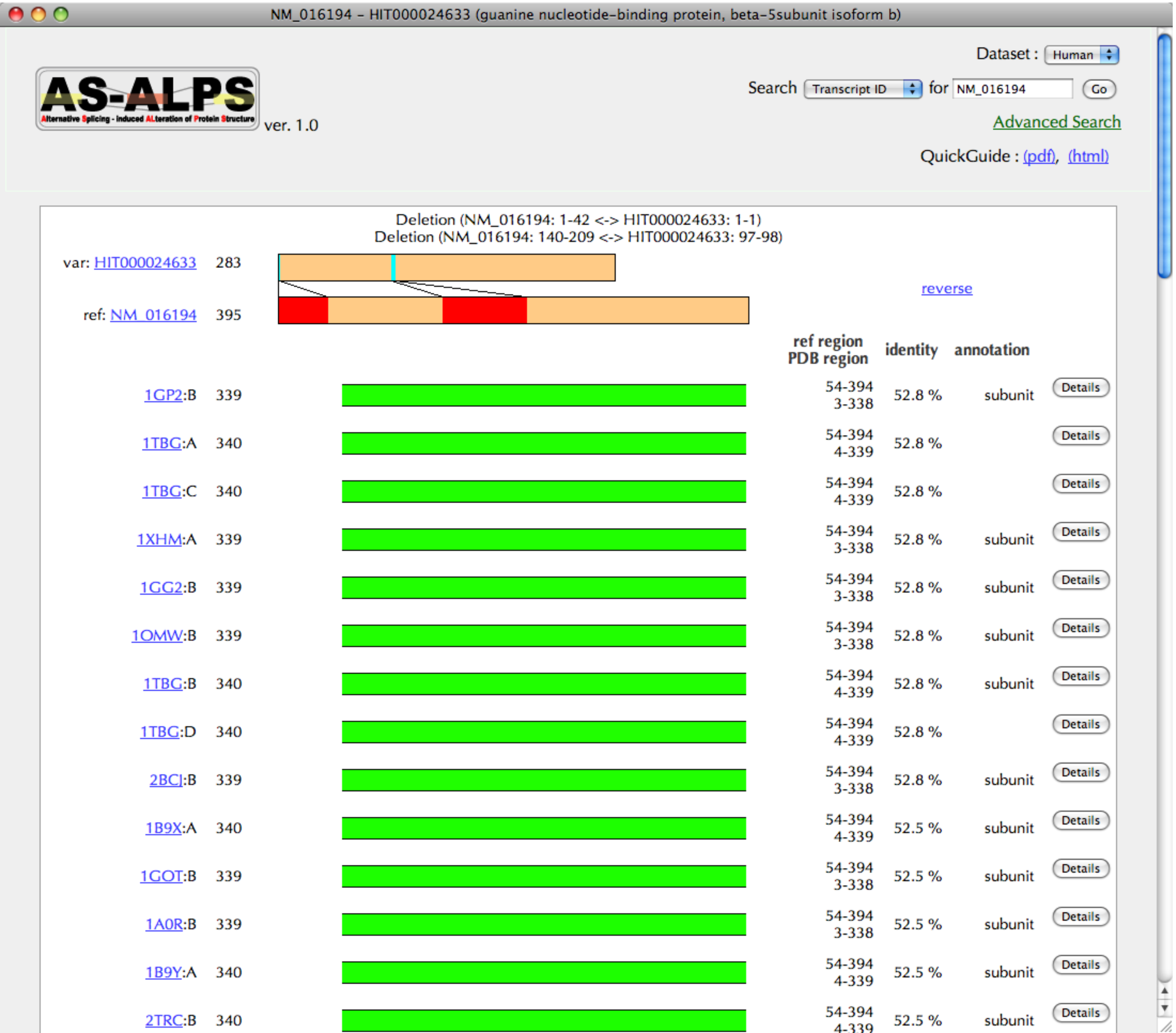
You can see here AS regions detected by a comparison between a pre-defined reference transcript and one variant transcript selected in the previous page.

For the substitution type, identity between amino acid sequences of AS regions mutually replaced is shown if comparable.

Click the button to show results of 3D structure assignment to AS regions.

→ Go to Page 7

Assigned 3D Structures Page



NM_016194 – HIT000024633 (guanine nucleotide-binding protein, beta-5subunit isoform b)

Deletion (NM_016194: 1-42 <-> HIT000024633: 1-1)
Deletion (NM_016194: 140-209 <-> HIT000024633: 97-98)

var: HIT000024633 283

ref: NM_016194 395

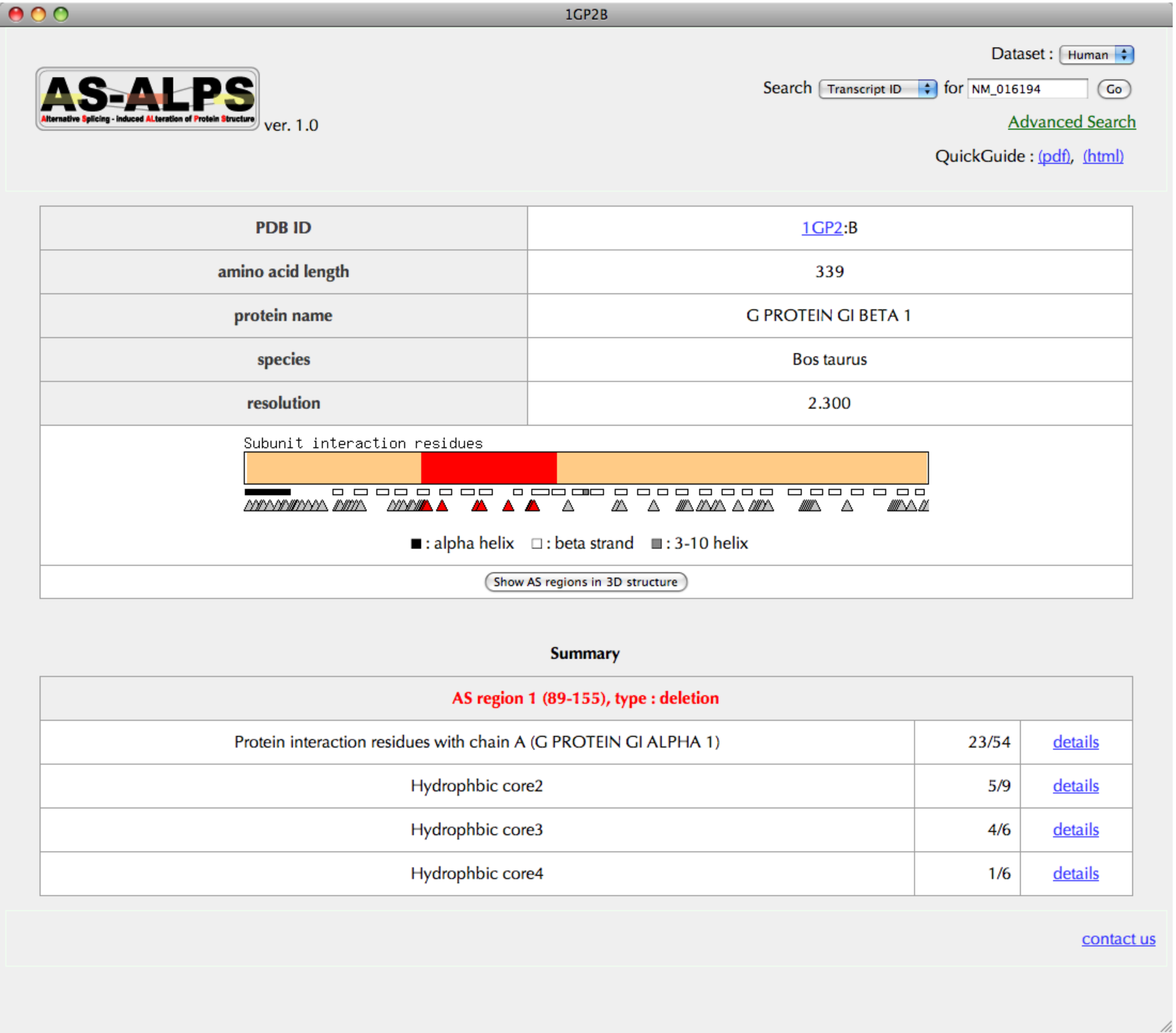
		ref region PDB region	identity	annotation	
1GP2:B	339	54-394 3-338	52.8 %	subunit	Details
1TBG:A	340	54-394 4-339	52.8 %		Details
1TBG:C	340	54-394 4-339	52.8 %		Details
1XHM:A	339	54-394 3-338	52.8 %	subunit	Details
1GG2:B	339	54-394 3-338	52.8 %	subunit	Details
1OMW:B	339	54-394 3-338	52.8 %	subunit	Details
1TBG:B	340	54-394 4-339	52.8 %	subunit	Details
1TBG:D	340	54-394 4-339	52.8 %		Details
2BCJ:B	339	54-394 3-338	52.8 %	subunit	Details
1B9X:A	340	54-394 4-339	52.5 %	subunit	Details
1GOT:B	339	54-394 3-338	52.5 %	subunit	Details
1A0R:B	339	54-394 3-338	52.5 %	subunit	Details
1B9Y:A	340	54-394 4-339	52.5 %	subunit	Details
2TRC:B	340	54-394 4-339	52.5 %	subunit	Details

You can see here a schematic representation of alignments between a pair of transcripts and PDB entries with shown sequence identities to the reference transcript.

You can also check in the "annotation" column, whether each PDB entry includes interaction info.

Click a button in the rightmost column to see details of interaction info.
→ Go to Page 8

3D Structure-related Interaction Info Page



PDB ID	1GP2:B
amino acid length	339
protein name	G PROTEIN GI BETA 1
species	Bos taurus
resolution	2.300

■ : alpha helix □ : beta strand ▨ : 3-10 helix

Summary

AS region 1 (89-155), type : deletion		
Protein interaction residues with chain A (G PROTEIN GI ALPHA 1)	23/54	details
Hydrophbic core2	5/9	details
Hydrophbic core3	4/6	details
Hydrophbic core4	1/6	details

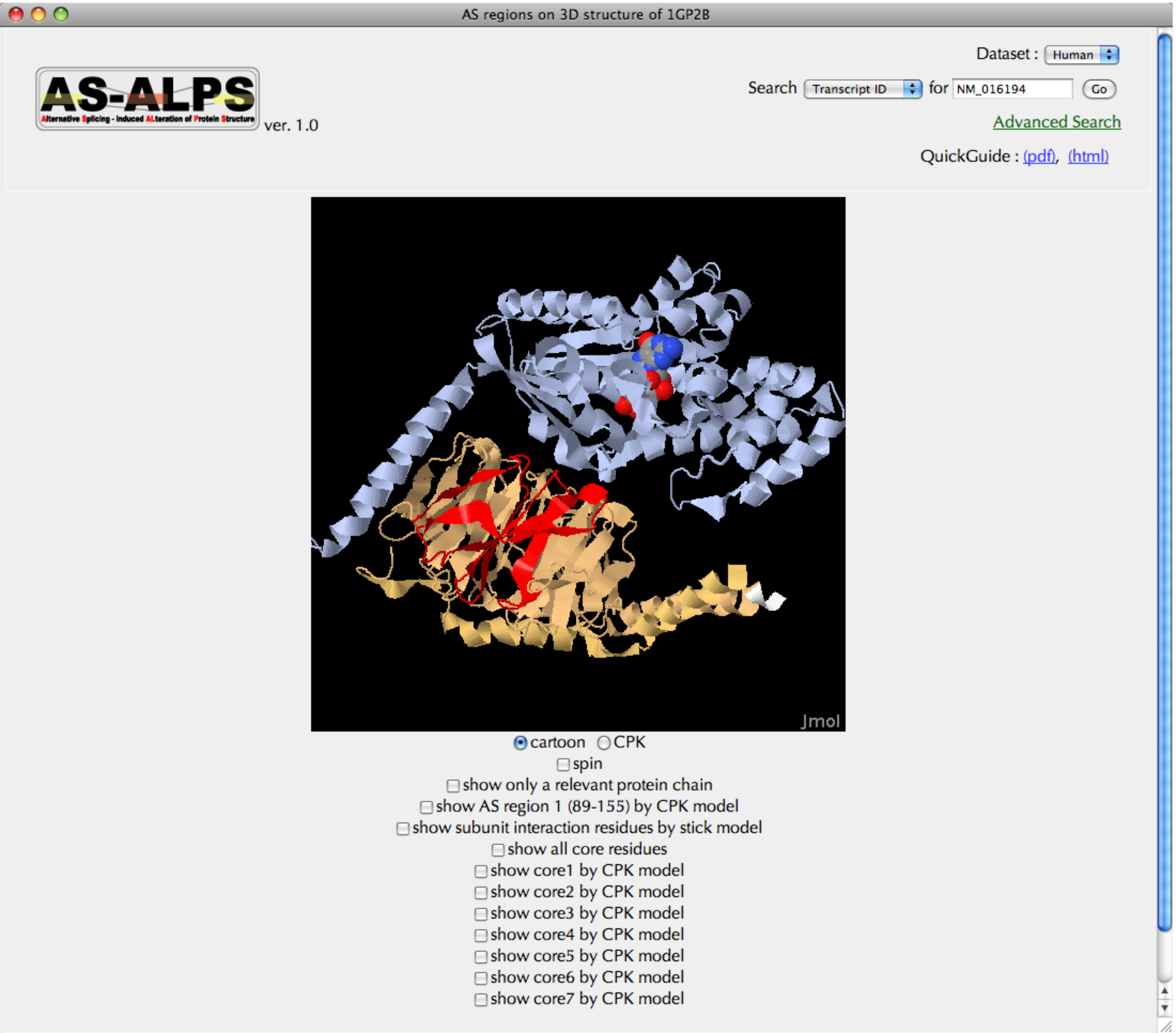
Interaction sites are depicted with triangles below a bar, in which AS regions are coloured by their types (deletion: red, insertion: cyan, substitution: violet) and constitutive regions are in light orange.

Click the button to show AS regions in 3D structure.
→ Go to Page 9

The lower half of the page summarizes the number of residues relating to interaction and hydrophobic core.

Details of interaction sites
→ Go to Page10
Details of core residues
→ Go to Page 11

AS regions on 3D Structure Page



AS regions on 3D structure of 1GP2B

ver. 1.0

Dataset : Human

Search Transcript ID for NM_016194 Go

Advanced Search

QuickGuide : (pdf), (html)

● cartoon ○ CPK

☐ spin

☐ show only a relevant protein chain

☐ show AS region 1 (89-155) by CPK model

☐ show subunit interaction residues by stick model

☐ show all core residues

☐ show core1 by CPK model

☐ show core2 by CPK model

☐ show core3 by CPK model

☐ show core4 by CPK model

☐ show core5 by CPK model

☐ show core6 by CPK model

☐ show core7 by CPK model

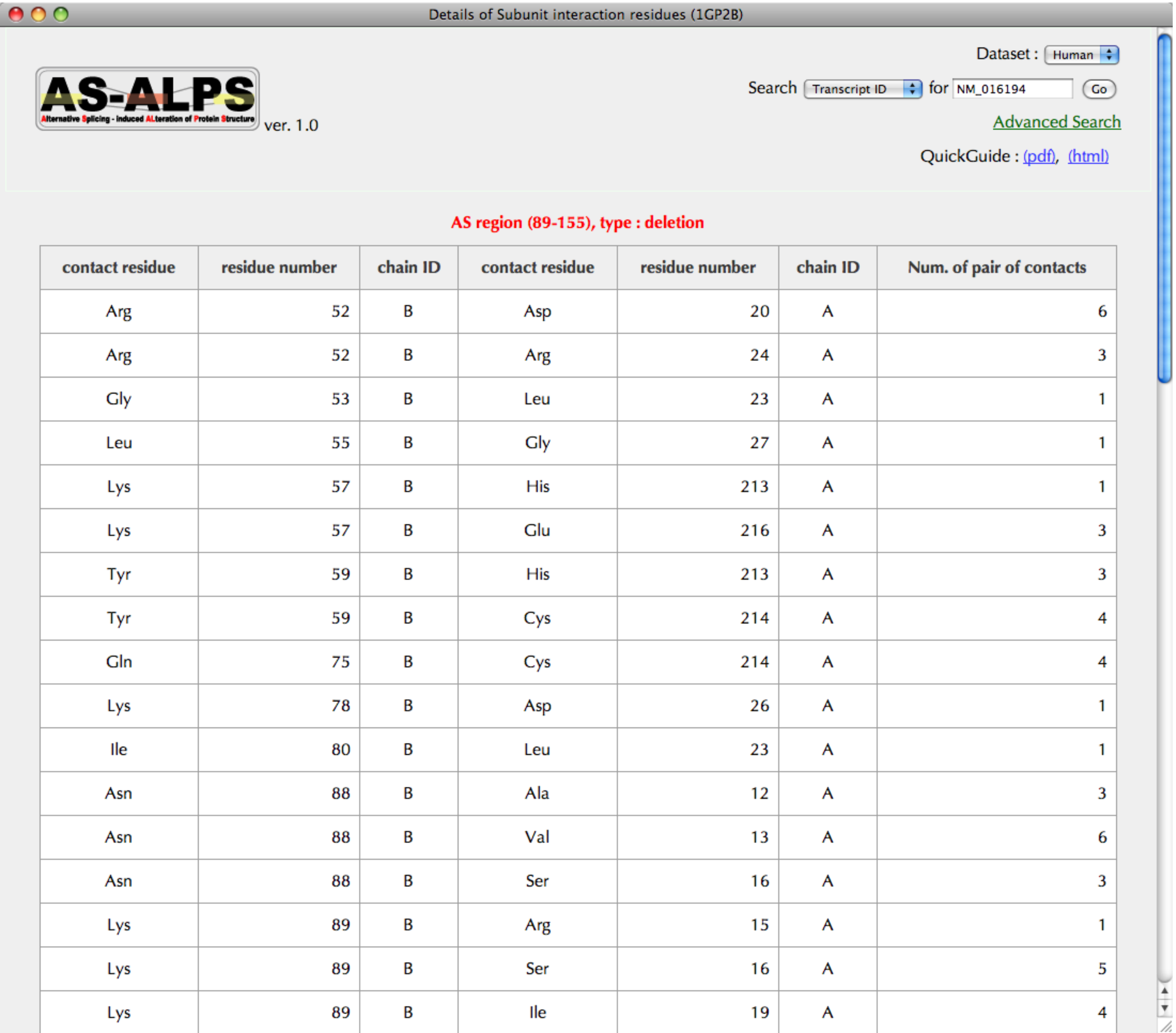
A relevant protein chain with AS regions is shown in the same colours as in page 8.

Other interacting proteins and DNA/RNAs, if any, are also shown in different colours.

You can check locations of hydrophobic cores by clicking check boxes below the 3D viewer.

Details of interaction sites extracted from 3D structure

Residues included in a relevant AS region are coloured by its type (deletion: red, insertion: cyan, substitution: violet).



Dataset : Human

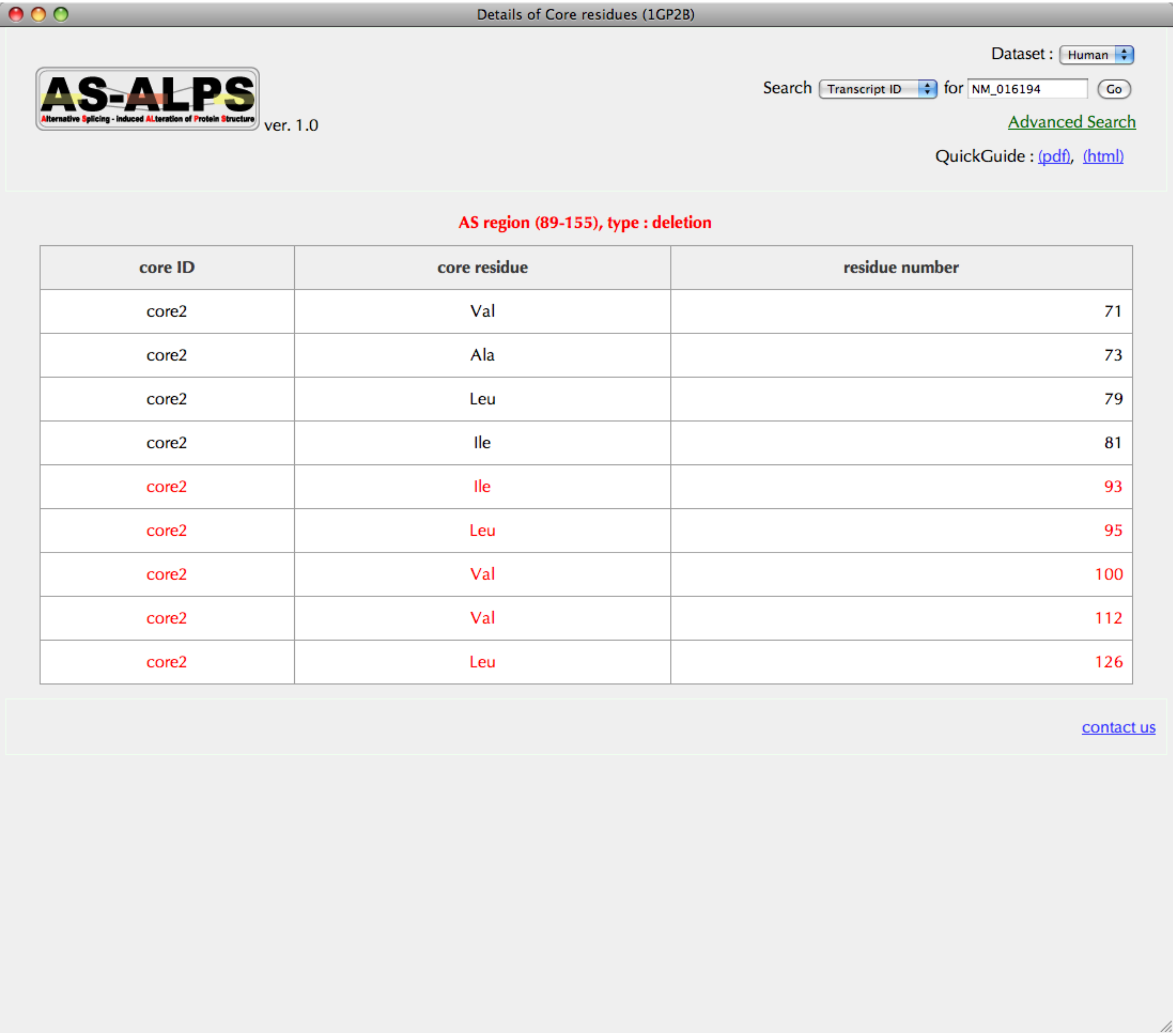
Search Transcript ID for NM_016194

QuickGuide : (pdf), (html)

AS region (89-155), type : deletion

contact residue	residue number	chain ID	contact residue	residue number	chain ID	Num. of pair of contacts
Arg	52	B	Asp	20	A	6
Arg	52	B	Arg	24	A	3
Gly	53	B	Leu	23	A	1
Leu	55	B	Gly	27	A	1
Lys	57	B	His	213	A	1
Lys	57	B	Glu	216	A	3
Tyr	59	B	His	213	A	3
Tyr	59	B	Cys	214	A	4
Gln	75	B	Cys	214	A	4
Lys	78	B	Asp	26	A	1
Ile	80	B	Leu	23	A	1
Asn	88	B	Ala	12	A	3
Asn	88	B	Val	13	A	6
Asn	88	B	Ser	16	A	3
Lys	89	B	Arg	15	A	1
Lys	89	B	Ser	16	A	5
Lys	89	B	Ile	19	A	4

Details of hydrophobic core residues extracted from 3D structure



Dataset : Human

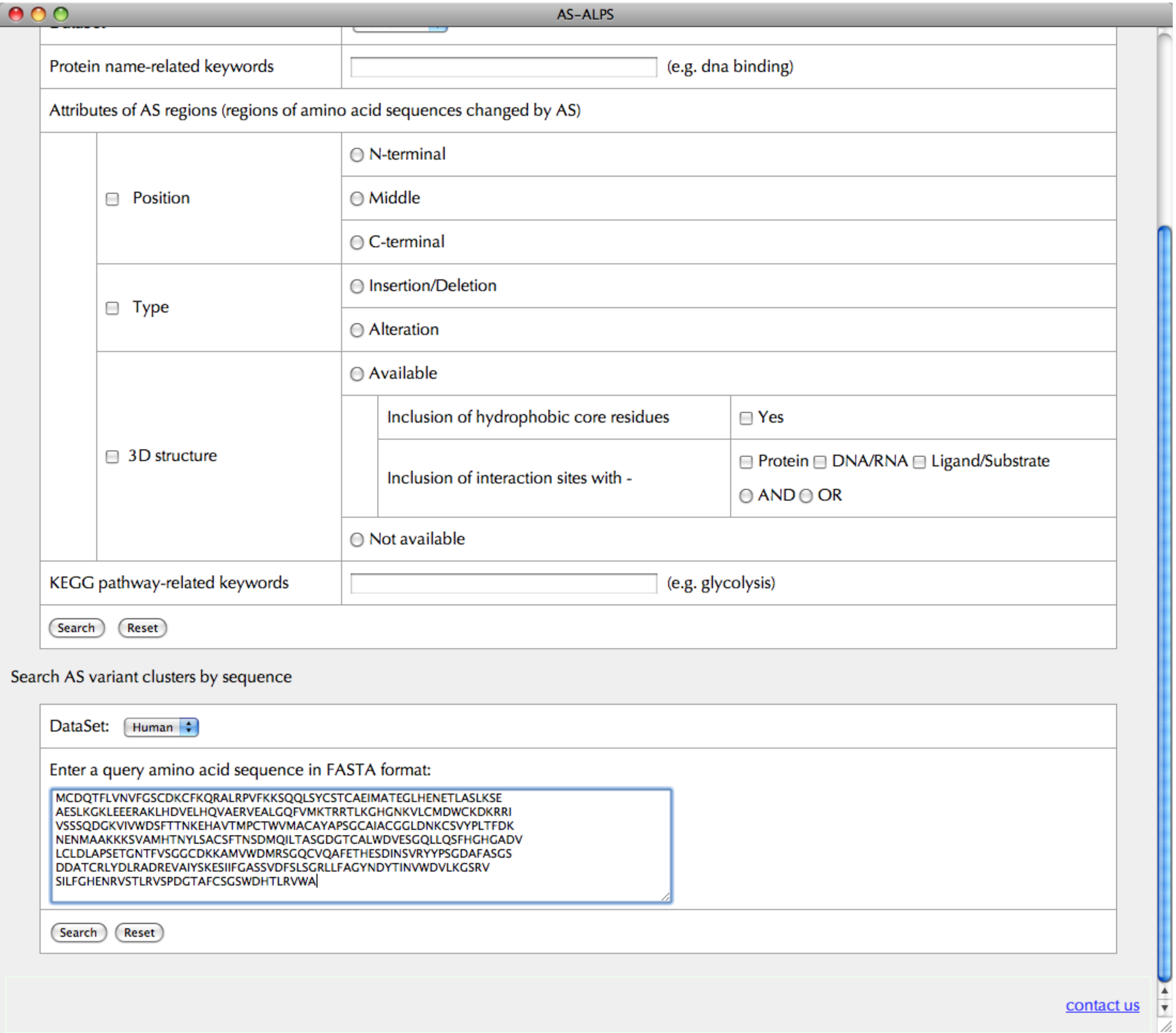
Search Transcript ID for NM_016194

AS region (89-155), type : deletion

core ID	core residue	residue number
core2	Val	71
core2	Ala	73
core2	Leu	79
core2	Ile	81
core2	Ile	93
core2	Leu	95
core2	Val	100
core2	Val	112
core2	Leu	126

Residues included in a relevant AS region are coloured by its type (deletion: red, insertion: cyan, substitution: violet).

Advanced Search Page 2



Protein name-related keywords	(e.g. dna binding)	
Attributes of AS regions (regions of amino acid sequences changed by AS)		
☐ Position	○ N-terminal ○ Middle ○ C-terminal	
☐ Type	○ Insertion/Deletion ○ Alteration	
☐ 3D structure	○ Available	
	Inclusion of hydrophobic core residues	☐ Yes
	Inclusion of interaction sites with -	☐ Protein ☐ DNA/RNA ☐ Ligand/Substrate ○ AND ○ OR
	○ Not available	
KEGG pathway-related keywords	(e.g. glycolysis)	

Search AS variant clusters by sequence

DataSet: Human

Enter a query amino acid sequence in FASTA format:

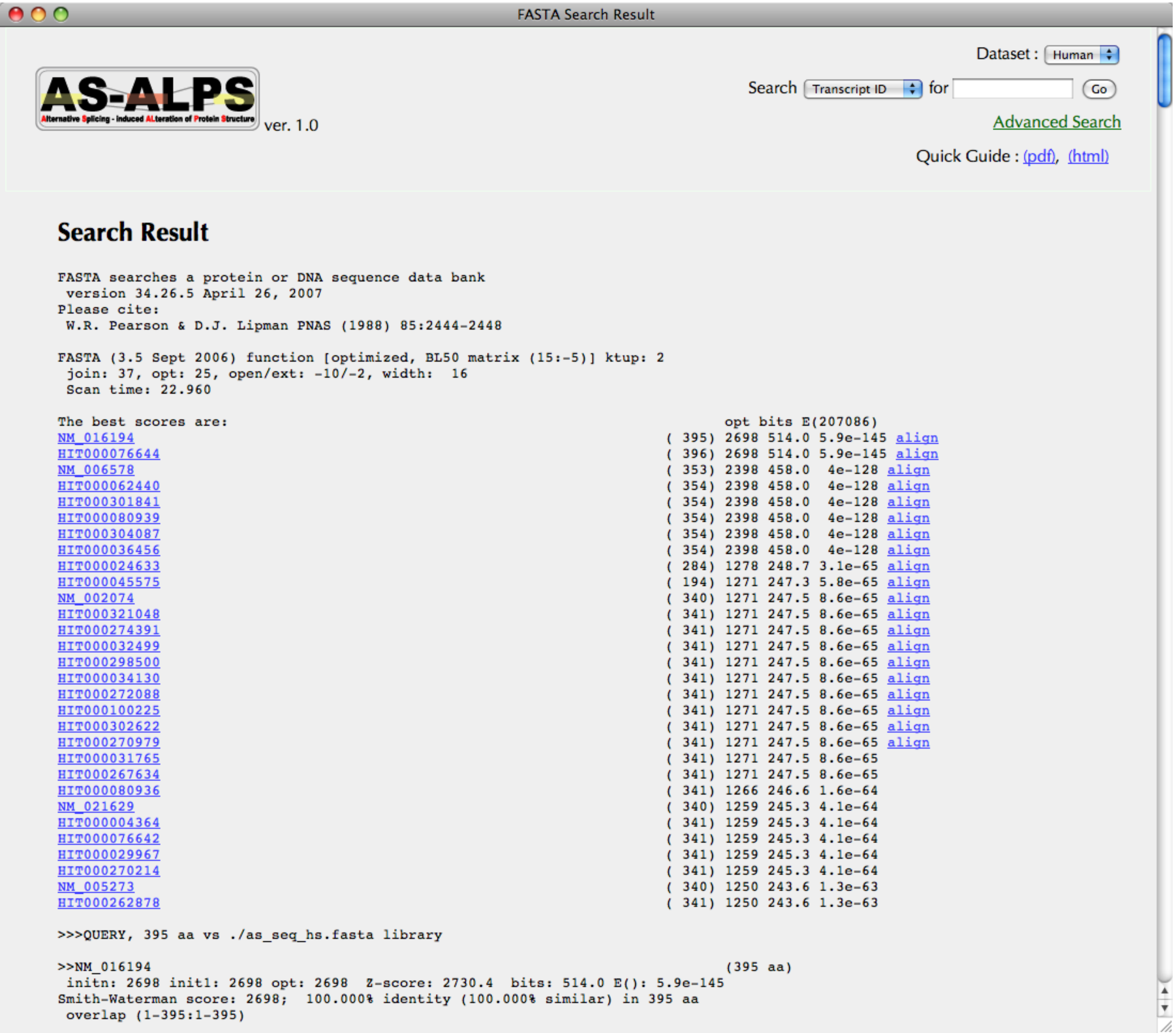
```
MCDQTFLVNVFGSCDKCFKQRALRPVFKKSQQLSYCSTCAEIMATEGLHENETLASLKSE
AESLKGKLEEERAKLHDVELHQVAERVEALGQFVMKTRRTLKGHGNKVLCMDWCKDKRRI
VSSSQDGKVIVWDSFTTNKEHAVTMPCTWVMACAYAPSGCAIACGGLDNKCSVYPLTFDK
NENMAAKKKSVAMHTNYLSACSFTNSDMQILTASGDGTCALWDVESGQLLQSFHGHGADV
LCLDLAPSETGNTFVSGGCDKKAMVWDMRSGQCVQAFETHESDINSVRYYPSGDAFASGS
DDATCRLYDLRADREVAIYSKESIIFGASSVDFSLSGRLLFAGYNDYTINVWDVLKGSRV
SILFGHENRVSTLRVSPDGTAFCSGSWDHTLRVWA
```

Search by sequence

You can search entries in this database by amino acid sequence. The search engine is FASTA.

→ Go to Page 13

FASTA Search Result Page 1



Dataset : Human

Search Transcript ID for

Search Result

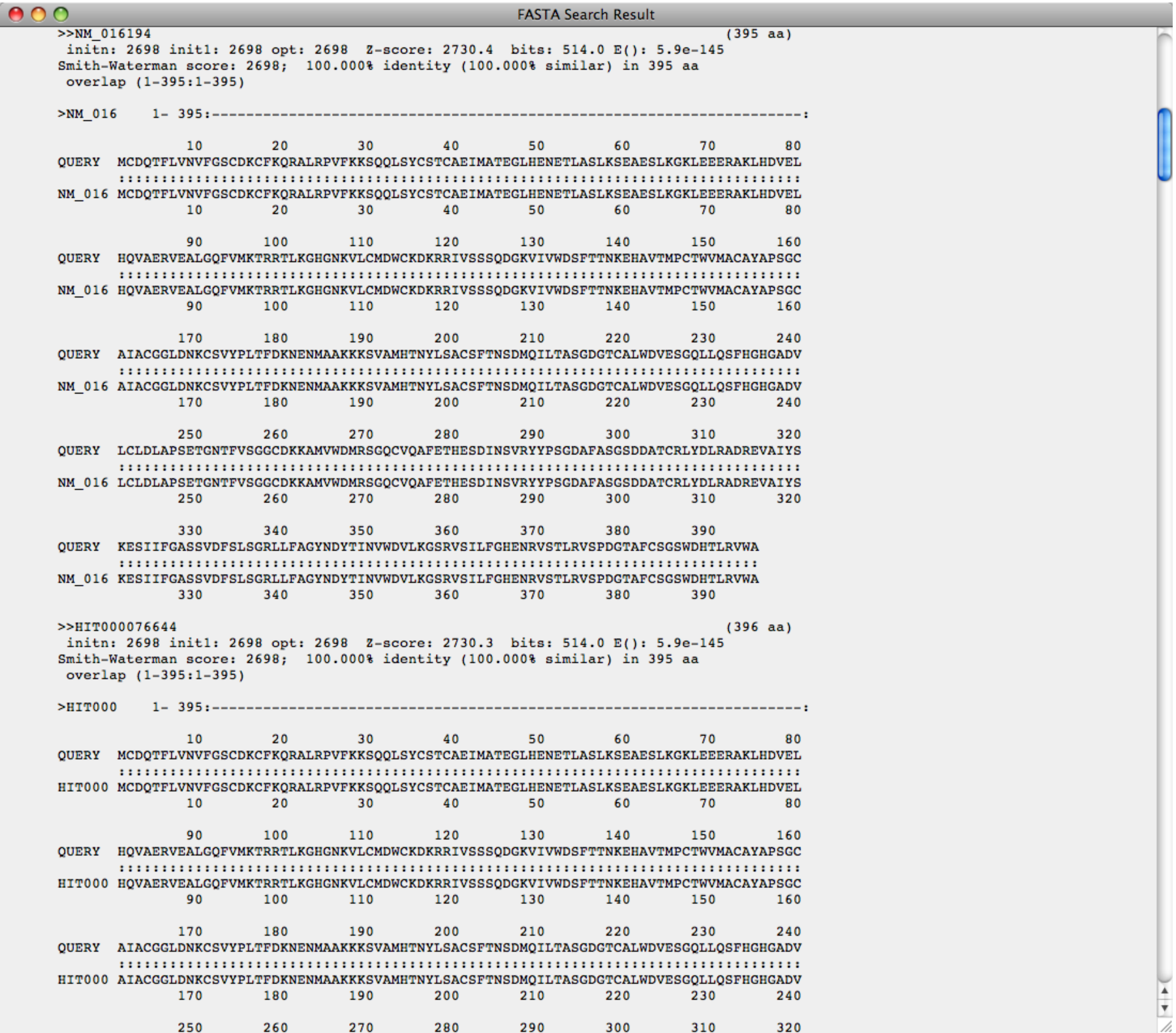
```
FASTA searches a protein or DNA sequence data bank
 version 34.26.5 April 26, 2007
Please cite:
 W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

FASTA (3.5 Sept 2006) function [optimized, BL50 matrix (15:-5)] ktup: 2
 join: 37, opt: 25, open/ext: -10/-2, width:  16
 Scan time: 22.960

The best scores are:                                                          opt bits E(207086)
NM_016194                                                              ( 395) 2698 514.0 5.9e-145 align
HIT000076644                                                           ( 396) 2698 514.0 5.9e-145 align
NM_006578                                                              ( 353) 2398 458.0  4e-128 align
HIT000062440                                                           ( 354) 2398 458.0  4e-128 align
HIT000301841                                                           ( 354) 2398 458.0  4e-128 align
HIT000080939                                                           ( 354) 2398 458.0  4e-128 align
HIT000304087                                                           ( 354) 2398 458.0  4e-128 align
HIT000036456                                                           ( 354) 2398 458.0  4e-128 align
HIT000024633                                                           ( 284) 1278 248.7 3.1e-65 align
HIT000045575                                                           ( 194) 1271 247.3 5.8e-65 align
NM_002074                                                              ( 340) 1271 247.5 8.6e-65 align
HIT000321048                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000274391                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000032499                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000298500                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000034130                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000272088                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000100225                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000302622                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000270979                                                           ( 341) 1271 247.5 8.6e-65 align
HIT000031765                                                           ( 341) 1271 247.5 8.6e-65
HIT000267634                                                           ( 341) 1271 247.5 8.6e-65
HIT000080936                                                           ( 341) 1266 246.6 1.6e-64
NM_021629                                                              ( 340) 1259 245.3 4.1e-64
HIT000004364                                                           ( 341) 1259 245.3 4.1e-64
HIT000076642                                                           ( 341) 1259 245.3 4.1e-64
HIT000029967                                                           ( 341) 1259 245.3 4.1e-64
HIT000270214                                                           ( 341) 1259 245.3 4.1e-64
NM_005273                                                              ( 340) 1250 243.6 1.3e-63
HIT000262878                                                           ( 341) 1250 243.6 1.3e-63

>>>QUERY, 395 aa vs ./as_seq_hs.fasta library

>>NM_016194                                                                   (395 aa)
 initn: 2698 init1: 2698 opt: 2698  Z-score: 2730.4  bits: 514.0 E(): 5.9e-145
Smith-Waterman score: 2698;  100.000% identity (100.000% similar) in 395 aa
 overlap (1-395:1-395)
```

You can see here a list of transcripts found by FASTA search.

Click one of transcript IDs in the leftmost column.
→ Go to Page 3

If you would like to check sequence alignments, click ”align” in the rightmost column.
→ Go to Page 14

FASTA Search Result Page 2



```
>>NM_016194                                                                   (395 aa)
 initn: 2698 init1: 2698 opt: 2698  Z-score: 2730.4  bits: 514.0 E(): 5.9e-145
Smith-Waterman score: 2698;  100.000% identity (100.000% similar) in 395 aa
 overlap (1-395:1-395)

>NM_016    1- 395:---------------------------------------------------------------------:

               10        20        30        40        50        60        70        80
QUERY  MCDQTFLVNVFGSCDKCFKQRALRPVFKKSQQLSYCSTCAEIMATEGLHENETLASLKSEAESLKGKLEEERAKLHDVEL
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
NM_016 MCDQTFLVNVFGSCDKCFKQRALRPVFKKSQQLSYCSTCAEIMATEGLHENETLASLKSEAESLKGKLEEERAKLHDVEL
               10        20        30        40        50        60        70        80

               90       100       110       120       130       140       150       160
QUERY  HQVAERVEALGQFVMKTRRTLKGHGNKVLCMDWCKDKRRIVSSSQDGKVIVWDSFTTNKEHAVTMPCTWVMACAYAPSGC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
NM_016 HQVAERVEALGQFVMKTRRTLKGHGNKVLCMDWCKDKRRIVSSSQDGKVIVWDSFTTNKEHAVTMPCTWVMACAYAPSGC
               90       100       110       120       130       140       150       160

              170       180       190       200       210       220       230       240
QUERY  AIACGGLDNKCSVYPLTFDKNENMAAKKKSVAMHTNYLSACSFTNSDMQILTASGDGTCALWDVESGQLLQSFHGHGADV
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
NM_016 AIACGGLDNKCSVYPLTFDKNENMAAKKKSVAMHTNYLSACSFTNSDMQILTASGDGTCALWDVESGQLLQSFHGHGADV
              170       180       190       200       210       220       230       240

              250       260       270       280       290       300       310       320
QUERY  LCLDLAPSETGNTFVSGGCDKKAMVWDMRSGQCVQAFETHESDINSVRYYPSGDAFASGSDDATCRLYDLRADREVAIYS
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
NM_016 LCLDLAPSETGNTFVSGGCDKKAMVWDMRSGQCVQAFETHESDINSVRYYPSGDAFASGSDDATCRLYDLRADREVAIYS
              250       260       270       280       290       300       310       320

              330       340       350       360       370       380       390
QUERY  KESIIFGASSVDFSLSGRLLFAGYNDYTINVWDVLKGSRVSILFGHENRVSTLRVSPDGTAFCSGSWDHTLRVWA
       :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
NM_016 KESIIFGASSVDFSLSGRLLFAGYNDYTINVWDVLKGSRVSILFGHENRVSTLRVSPDGTAFCSGSWDHTLRVWA
              330       340       350       360       370       380       390

>>HIT000076644                                                                 (396 aa)
 initn: 2698 init1: 2698 opt: 2698  Z-score: 2730.3  bits: 514.0 E(): 5.9e-145
Smith-Waterman score: 2698;  100.000% identity (100.000% similar) in 395 aa
 overlap (1-395:1-395)

>HIT000    1- 395:---------------------------------------------------------------------:

               10        20        30        40        50        60        70        80
QUERY  MCDQTFLVNVFGSCDKCFKQRALRPVFKKSQQLSYCSTCAEIMATEGLHENETLASLKSEAESLKGKLEEERAKLHDVEL
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
HIT000 MCDQTFLVNVFGSCDKCFKQRALRPVFKKSQQLSYCSTCAEIMATEGLHENETLASLKSEAESLKGKLEEERAKLHDVEL
               10        20        30        40        50        60        70        80

               90       100       110       120       130       140       150       160
QUERY  HQVAERVEALGQFVMKTRRTLKGHGNKVLCMDWCKDKRRIVSSSQDGKVIVWDSFTTNKEHAVTMPCTWVMACAYAPSGC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
HIT000 HQVAERVEALGQFVMKTRRTLKGHGNKVLCMDWCKDKRRIVSSSQDGKVIVWDSFTTNKEHAVTMPCTWVMACAYAPSGC
               90       100       110       120       130       140       150       160

              170       180       190       200       210       220       230       240
QUERY  AIACGGLDNKCSVYPLTFDKNENMAAKKKSVAMHTNYLSACSFTNSDMQILTASGDGTCALWDVESGQLLQSFHGHGADV
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
HIT000 AIACGGLDNKCSVYPLTFDKNENMAAKKKSVAMHTNYLSACSFTNSDMQILTASGDGTCALWDVESGQLLQSFHGHGADV
              170       180       190       200       210       220       230       240

              250       260       270       280       290       300       310       320
```

You can see here an alignment between your query sequence and a sequence picked up from this database by FASTA search.