

WORKED EXAMPLE – BROWSER (e76)

In this worked example we will explore the human **ABCD1** (ATP-binding cassette, sub-family D (ALD), member 1) gene.

The protein encoded by this gene is likely involved in the peroxisomal transport or catabolism of very long chain fatty acids (VLCFAs). Mutations in the **ABCD1** gene can cause **Adrenoleukodystrophy**, a rare X-linked disorder that causes a range of clinical phenotypes, often leading to a vegetative state and/or death (see also <http://en.wikipedia.org/wiki/Adrenoleukodystrophy>).

☞ Go to the Ensembl homepage (<http://www.ensembl.org/>).

The screenshot shows the Ensembl genome browser homepage. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, and More. A search bar is located on the right. Below the navigation bar, there is a main content area with several sections. On the left, there is a 'Browse a Genome' section with a search bar and a 'Go' button. Below this, there are 'Popular genomes' listed: Human (GRCh38), Mouse (GRCm38), and Zebrafish (Zv9). There is also a 'Log in to customize this list' link and a 'Select a species' dropdown menu. In the center, there are several boxes for 'ENCODE data in Ensembl', 'Variant Effect Predictor', 'Gene expression in different tissues', 'Find SNPs and other variants for my gene', 'Retrieve gene sequence', and 'Compare genes across species'. On the right, there is a 'What's New in Release 76 (August 2014)' section with a list of updates, including 'Updated human assembly to GRCh38', 'New BLAST/BLAT interface', 'New regulation displays', and 'New species: Amazon molly and Olive baboon'. Below this, there are links for 'Full details of this release', 'All web updates by release', and 'More release news on our blog'. There is also a 'Latest blog posts' section with a list of recent posts. At the bottom right, there is a 'Did you know...?' section.

Searching

First of all, let's search for the human **ABCD1** gene.

☞ Select 'Search: Human' and type 'abcd1' in the 'for' text box.

☞ Click [Go].

The search result shows an **ABCD1** gene and several transcripts (splice variants).

Only searching Human

2071 results match **abcd1** when restricted to species: Human X

ABCD1 (Human Gene)

ENSG00000101986 X:153724868-153744762:1

ATP-binding cassette, sub-family D (ALD), member 1 [Source:HGNC Symbol;Acc:HGNC:61] **ABCD1**
(Vega gene) is associated with Gene ENSG00000101986

[Variation table](#) • [Location](#) • [Regulation](#) • [Orthologues](#) • [Gene tree](#)

ABCD1-001 (Human Transcript)

ENST00000218104 X:153724868-153744762:1

ATP-binding cassette, sub-family D (ALD), member 1 [Source:HGNC Symbol;Acc:HGNC:61] **ABCD1-001**
(Vega transcript) is associated with Transcript ENST00000218104

[Location](#) • [cDNA seq.](#) • [Variation table](#) • [Protein seq.](#) • [Population](#) • [Protein](#)

ABCD1-003 (Human Transcript)

ENST00000370129 X:153725817-153729897:1

ATP-binding cassette, sub-family D (ALD), member 1 [Source:HGNC Symbol;Acc:HGNC:61] **ABCD1-003**
(Vega transcript) is associated with Transcript ENST00000370129

[Location](#) • [cDNA seq.](#) • [Variation table](#) • [Protein seq.](#) • [Population](#) • [Protein](#)

Click on 'ABCD1 (Human Gene)' (the first hit)

This leads us to the 'Gene summary' page under the 'Gene' tab.

The Gene tab

Pages (also called 'views') in Ensembl are organised under a number of tabs, i.e. 'Species', 'Location', 'Gene', 'Transcript', 'Variation' and 'Regulation'. The various available pages under each tab are listed in the left-hand side menu.

The 'Gene Summary' page shows general information about the *ABCD1* gene and the transcripts that have been annotated for it as part of the GENCODE gene set (<http://www.genencodegenes.org/>). Note the information icon (i) next to 'Gene summary' that opens up a help page, as well as the legend at the bottom of the graphical display.

Click [Show transcript table].

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | More ▾ Login/Register

Human (GRCh38) ▾ Location: X:153,724,868-153,744,762 Gene: ABCD1

Gene-based displays

- Summary
- Splice variants (3)
- Transcript comparison
- Supporting evidence
- Sequence
 - Secondary Structure
- External references
- Regulation
- Expression
- Comparative Genomics
 - Genomic alignments
 - Gene tree (image)
 - Gene tree (text)
 - Gene tree (alignment)
 - Gene gain/loss tree
- Orthologues (55)
- Paralogues (3)
- Protein families (3)
- Phenotype
- Genetic Variation

Gene: ABCD1 ENSG00000101986

Description ATP-binding cassette, sub-family D (ALD), member 1 [Source:HGNC Symbol;Acc:HGNC:61]

Synonyms adrenoleukodystrophy, ALD, ALDP, AMN

Location [Chromosome X: 153,724,868-153,744,762](#) forward strand.

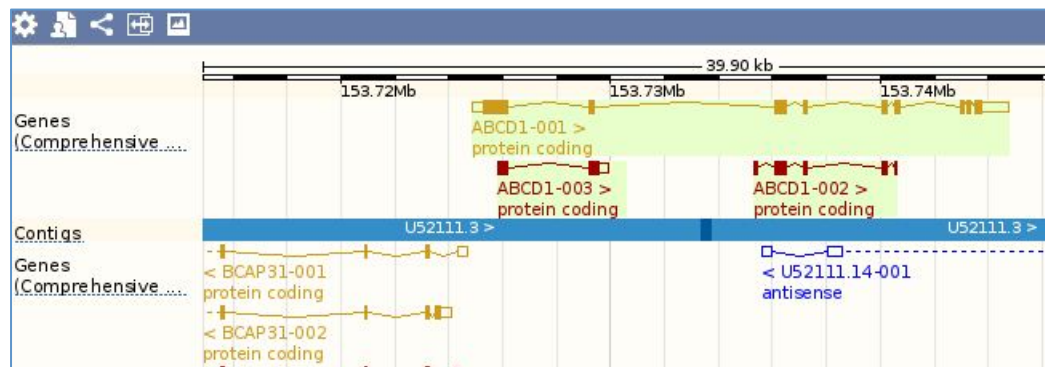
INSDC coordinates chromosome:GRCh38:CM000685.2:153724868:153744762:1

Transcripts This gene has 3 transcripts (splice variants) Hide transcript table

Show/hide columns (1 hidden)							Filter
Name	Transcript ID	Length	Protein	Biotype	CCDS	RefSeq	Flags
ABCD1-001	ENST00000218104	3664 bp	745 aa (view)	Protein coding	CCDS14728	NM_000033 NP_000024	GENCODE basic
ABCD1-003	ENST00000370129	1016 bp	227 aa (view)	Protein coding	-	-	GENCODE basic
ABCD1-002	ENST00000443684	668 bp	223 aa (view)	Protein coding	-	-	CDS 5' and 3' incomplete

You can customise the table by clicking on 'Show/hide columns'. For example, turn on the UniProt matches, and turn off the Flags.

Looking in the table, ABCD1-001 (ENST00000218104) has a CCDS (Consensus Coding Sequence; <http://www.ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi>) identifier. The CCDS is a set of coding sequences about which Ensembl, Havana, NCBI and UCSC all agree and which therefore are considered to be of high quality. In the transcript table it can be seen that two transcripts are part of the GENCODE Basic Set, which is intended to provide a simplified subset of the GENCODE transcript annotations. Transcript ABCD1-002 / ENST00000443684 is not part of this set, because its CDS is incomplete.



The graphical display shows the same three transcripts as the table. You can click on a transcript to learn more about it.

Protein-coding transcripts are gold or red. Gold transcripts are identical between the Ensembl and Havana projects, thus reflect a high standard. Red transcripts have either been annotated by Ensembl or Havana. In this case, ABCD1-002 and ABCD1-003 have been annotated by Havana.

Boxes and lines in the transcripts represent exons and introns, respectively. Empty boxes represent untranslated regions (UTRs), while filled boxes represent the coding sequence (CDS).

Summary – ABCD1 Transcripts

- There are three transcripts, all protein coding
- ABCD1-001 is gold, a symbol of high quality

The *ABCD1* gene is located on the forward strand of the genome. This can be seen from the arrows next to the transcript names, which indicate the direction of transcription and from the fact that the transcript models are shown above the blue bar that represents the genome. Transcripts located on the reverse strand are shown below the blue bar.

☞ Click on 'Sequence' in the side menu.

Marked-up sequence

 Download sequence  BLAST this sequence

Key

Exons

```
>chromosome:GRCh38:X:153724268:153745362:1
CCTCGTCGATGGGCGGGGAGCCTCCGCGGTCCCGGAGCCAGCCGCGCGCGGAGCCC
GCTCACCAGAGTTTCCACAGTCAACGTGCAGGCCCGCCGCGAGCAACAGAACTCTCCAC
AGCAGCCCCCGGCCCGCCCTCATACCGCGGCCGGAACCGGAAGCGCCCGCGGGCACC
GCCCACAGCCCTCGCGAGGCCCGGAGGCTCCGCCACCTCGCTTCCCACCCGCCCC
GGAGCGGAGGGCCGCGCTCCGAGCGGGAGAGGAAGAGGCGCCTCGGCTCCGGGCGAGC
AGGGCGGGGTGGAGCGAGCAGCGGGCGGGCGGGCGGGGCTTTGTGCGGCGGGCGAGG
GCCGCTTCTCTAGTCCGCGCGGCCGTCCACGTCTCTGTGGTGGCGGAGGGGCCCGCG
AGGGCGGAGAACGGGAGGTGGGGGTGTGGGCGGGCCCCCGGAGGGGCGAGAACAGGGTG
GGGCTCCGCGCCCGGACTCCGCCCTCGCCCCCTCCTCCGCTCCTCCCTTCCCCGAC
TCGCCCCCTGGGGAAGAGTGGGTGGGATTCTGGGCCGTGGAGGAGTCACTGTCGCTTCA
GCCAGGCTGCGGAGCGGACGGACGCGCTGGTGCCCCGGGGAGGGGCGCCACCGGGGAG
GAGGAGGAGGAGAAGGTGGAGAGGAAGAGACGCCCTCTGCCGAGACCTCTCAAGGCC
CTGACCTCAGGGGCCAGGGCACTGACAGGACAGGAGAGCAAGTTCTTCACTTGGGCTG
CCCGAAGAGGCCGCGACCCTGGAGGGCCCTGAGCCACCGCACAGGGGCCCCAGCACCA
CCCCGGGGGCTAAAGCGACAGTCTCAGGGGCCATCGCAAGGTTTCCAGTTGCCTAGACA
ACAGGCCCAGGGTCAGAGCAACAATCCTTCCAGCCACCTGCCTCAACTGCTGCCCCAGGC
ACCAGCCCCAGTCCCTACGCGGCAGCCAGCCAGGTGACATGCCGGTGCTCTCCAGGCCC
CGGCCCTGGCGGGGAACACGCTGAAGCGCACGGCCGTGCTCTGGCCCTCGCGGCCTAT
GGAGCCCACAAAGTCTACCCCTTGGTGCGCCAGTGCCCTGGCCCCGGCCAGGGGTCTTCAG
```

Exon from
neighbouring gene
(BCAP31)

ABCD1 Exon

On the 'Sequence' page the sequence of the *ABCD1* gene plus 600 bp upstream and downstream is shown. Exon sequences belonging to the *ABCD1* gene are shown in red letters on a peach background, while exons belonging to other genes are shown in black letters on a peach background. All possible exon sequence is shown, across all the transcripts.

Almost all graphical displays in Ensembl can be configured. This is done using the [Configure this page] button.

☞ Click [Configure this page] in the side menu.

A pop-up window lists all display options.

Summary – Gene Sequence

- All exons are highlighted, for all genes in the region
- ABCD1 exons are in bold letters

This opens the Variation tab for this sequence variant.

The screenshot shows the Variation tab for the SNP rs188990625. The top navigation bar includes 'Human (GRCh38)', 'Location: X:153,724,868-153,744,762', 'Gene: ABCD1', and 'Variation: rs188990625'. On the left, a 'Variation displays' sidebar lists various views like 'Genomic context', 'Genes and regulation (10)', 'Population genetics', etc. The main content area displays the variant details: 'rs188990625 SNP', 'Original source' (dbSNP), 'Alleles' (C/T), 'Location' (Chromosome X:153724783), 'Most severe consequence' (Upstream gene variant), 'Evidence status' (red X icon), and 'HGVS name' (X:g.153724783C>T). Below this is an 'Explore this variation' section with ten interactive icons: Genomic context, Genes and regulation, Population genetics, Individual genotypes, Linkage disequilibrium, Phenotype data, Citations, Phylogenetic context, and Flanking sequence.

This is a summary page showing basic information about the variant.

Summary – rs188990625

- It is a SNP with alleles C/T
- It's upstream of a gene
- It's found in the 1000 Genomes project

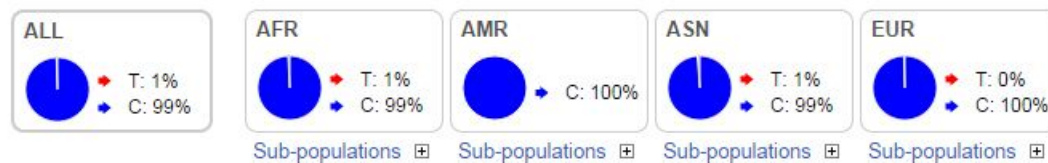
Either click on an icon or one of the left-hand links to navigate through the information for this SNP.

🖱️ Click on Population genetics

Find out more about the frequency of the T and C alleles across world populations.

Population genetics ⓘ

1000 Genomes Project Phase 1 allele frequencies



- **ALL**, All populations
- **AFR**, African
- **AMR**, Mixed American
- **ASN**, East Asian
- **EUR**, European

Expand Sub-populations for more of a break-down. This FAQ helps you interpret the population codes:

<http://www.ensembl.org/Help/Faq?id=328>

🔗 Click on 'Phenotype' in the side menu.

On the 'Phenotype' page phenotypes that have been associated with the *ABCD1* gene as well as with variants associated with the *ABCD1* gene are shown.

Phenotype ⓘ

List of phenotype(s) associated with the gene ENSG00000101986

Phenotype	Source	Locations
ADRENOLEUKODYSTROPHY	OMIMGENE	View on Karyotype
ADRENOMYELONEUROPATHY	Orphanet	View on Karyotype
X-linked cerebral adrenoleukodystrophy	Orphanet	View on Karyotype
Adrenoleukodystrophy, X-Linked	DDG2P	View on Karyotype

Phenotypes associated with the gene from variation annotations

Number of variants	Show/hide details	Phenotype	Locations	Biomart	Source(s)
486	Show	ALL variations with a phenotype annotation			-
1	Show	ADDISON DISEASE	View on Karyotype	-	OMIM
16	Show	ADRENOLEUKODYSTROPHY	View on Karyotype	-	OMIM
4	Show	ADRENOMYELONEUROPATHY	View on Karyotype	-	OMIM
407	Show	Annotated by HGMD but no phenotype description is publicly available	-	-	HGMD-PUBLIC

Now go back to information about the *ABCD1* gene.

☞ Click on the *Gene* tab

Human (GRCh38) ▼

Location: X:153,724,868-153,744,762

Gene: ABCD1

Variation: rs188990625

Gene-based displays

Summary

Gene: ABCD1 ENSG00000101986

We can see all variation for this gene.

☞ Click on 'Genetic Variation - Variation table' in the side menu.

On the 'Variation table' page all variants in *ABCD1* gene are shown, grouped by consequence type.

Summary of variation consequences in ENSG00000101986 [Switch to tree view](#)

Show All ▼ entries

Filter

Number of variant consequences	Type	Description
0 -	Transcript ablation	A feature ablation whereby the deleted region includes a transcript feature (SO:0001893)
31 Show	Splice donor variant	A splice variant that changes the 2 base region at the 5' end of an intron (SO:0001575)
31 Show	Splice acceptor variant	A splice variant that changes the 2 base region at the 3' end of an intron (SO:0001574)
108 Show	Stop gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript (SO:0001587)
198 Show	Frameshift variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three (SO:0001589)
1 Show	Stop lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript (SO:0001578)
4 Show	Initiator codon variant	A codon variant that changes at least one base of the first codon of a transcript (SO:0001582)
0 -	Transcript amplification	A feature amplification of a region containing a transcript (SO:0001889)
9 Show	Inframe insertion	An inframe non synonymous variant that inserts bases into in the coding sequence (SO:0001821)
42 Show	Inframe deletion	An inframe non synonymous variant that deletes bases from the coding sequence (SO:0001822)
621 Show	Missense variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved (SO:0001563)
48 Show	Splice region variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron (SO:0001630)
0 -	Incomplete terminal codon variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed (SO:0001826)
99 Show	Synonymous variant	A sequence variant where there is no resulting change to the encoded amino acid (SO:0001819)

☞ Click on 'Show' for 'Missense variant'.

This results in a list of all missense variants. Clicking on the ID of a variant will lead us to the 'Variation' tab, where more information about the variant in question can be found.

Show All entries															Show/hide columns					Filter		
ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin sig	Type	AA	AA co-ord	SI FT	Poly Phen	Transcript									
rs183021839	X:153725304	A/C	0.003 (C)	SNP	dbSNP		-	Missense variant	N/T	13	0.77	0	ENST00000218104									
X-ALD_c.38A>C	X:153725304	A/C	-	SNP	PhenCode	-	-	Missense variant	N/T	13	0.77	0	ENST00000218104									
X-ALD_c.55G>T	X:153725321	G/T	-	SNP	PhenCode	-	-	Missense variant	A/S	19	0.02	0.735	ENST00000218104									
rs375019683	X:153725366	C/T	-	SNP	dbSNP		-	Missense variant	P/S	34	0.01	0.947	ENST00000218104									

Clicking on a variant ID will lead us to the 'Variation' tab, where detailed information about the variant in question can be found.

Let's find out more about *ABCD1*.

☞ Click on External References in the side menu.

This shows matches to the Ensembl gene in other projects and databases. A table that links Ensembl transcripts to UniProt and RefSeq identifiers is found at the bottom of the page.

<div>The following database identifiers correspond to the transcripts of this gene:</div> <div>Filter</div>						
Transcript ID	CCDS	UniProtKB/ Swiss-Prot	RefSeq peptide	RefSeq mRNA	Vega transcript	UniProtKB/ TrEMBL
ENST00000218104	CCDS14728.1	P33897	NP_000024.2	NM_000033.3	OTTHUMT00000061041	
ENST00000370129					OTTHUMT00000061043	A6NEP8
ENST00000443684					OTTHUMT00000061042	

This is similar to what we saw in the Transcript table, but has more information.

☞ Click on 'Phenotype' in the side menu.

On the 'Phenotype' page phenotypes that have been associated with the *ABCD1* gene as well as with variants associated with the *ABCD1* gene are shown.

Phenotype

List of phenotype(s) associated with the gene ENSG00000101986

Filter 		
Phenotype	Source	Locations
Adrenoleukodystrophy, X-Linked	DDG2P	View on Karyotype
ADRENOLEUKODYSTROPHY	OMIMGENE	View on Karyotype
CADD	Orphanet	View on Karyotype
ADRENOMYELONEUROPATHY	Orphanet	View on Karyotype
X-linked cerebral adrenoleukodystrophy	Orphanet	View on Karyotype

Phenotypes associated with the gene from variation annotations

Show 10 entries Filter						
Number of variants	Show/hide details	Phenotype	Locations	Biomart	Source(s)	
475	Show	ALL variations with a phenotype annotation			-	
1	Show	ADRENOLEUKODYSTROPHY	View on Karyotype	-	OMIM	
1	Show	ADRENOMYELONEUROPATHY	View on Karyotype	-	OMIM	
410	Show	Annotated by HGMD but no phenotype description is publicly available	-	-	HGMD-PUBLIC	
1	Show	COSMIC:tumour_site:NS	View on Karyotype	View list in BioMart	COSMIC	
2	Show	COSMIC:tumour_site:autonomic_ganglia	View on Karyotype	View list in BioMart	COSMIC	
4	Show	COSMIC:tumour_site:breast	View on Karyotype	View list in BioMart	COSMIC	
1	Show	COSMIC:tumour_site:central_nervous_system	View on Karyotype	View list in BioMart	COSMIC	
19	Show	COSMIC:tumour_site:endometrium	View on Karyotype	View list in BioMart	COSMIC	
3	Show	COSMIC:tumour_site:haematopoietic_and_lymphoid_tissue	View on Karyotype	View list in BioMart	COSMIC	

🔗 Click on 'Comparative Genomics- Orthologues' in the side menu.

All *ABCD1* orthologues in other species as identified by Ensembl are shown in the 'Selected orthologues' table on the 'Orthologues' page. Table columns can be hidden using the [Show/hide columns] button. Rows can be (re)ordered using the triangles next to the column headers and filtered using the 'Filter' text box. The table can be exported as an Excel spreadsheet by clicking on the Excel icon.

🔗 Select 'Rodents' and find the mouse orthologue

Mouse (<i>Mus musculus</i>)	1-to-1	0.06621	ENSMUSG00000031378	<ul style="list-style-type: none"> Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image) 	X:73716597-73738534:1	90	89
Abcd1 ATP-binding cassette, sub-family D (ALD), member 1 [Source:MGI] Symbol;Acc:MGI:1349215							

For the human *ABCD1* gene, a 1-to-1 orthologue has been identified in mouse, i.e. ENSMUSG00000031378.

Summary – Human *ABCD1*

- Has three transcripts
- Has a match to UniProt P33897
- Has an orthologue in mouse
- Is linked to Adrenoleukodystrophy

Let's look specifically at one splice variant.

🔗 Click on 'ENST00000218104' in the transcript table at the top of the page.

This leads us to the 'Transcript summary' page under the 'Transcript' tab.

The Transcript tab

Note that, because we have moved from the 'Gene' tab to the 'Transcript' tab, the side menu has changed and now shows links to pages with information about this specific splice variant.

🔗 Click on 'Sequence - Exons' in the side menu.


On the 'Exons' page the sequence of the unspliced transcript is shown. The coding sequence (CDS) is shown in black, untranslated regions (UTRs) in purple, introns in blue and flanking sequences in green. By default only a small part of the introns and the flanking sequences is shown, but this can be changed on the configuration page.

Exons ⓘ

[Download sequence](#) [BLAST this sequence](#)

Key

Exons/Introns Flanking sequence Intron sequence Translated sequence
UTR

Show All ▾ entries		Show/hide columns				Filter	
No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence					gggaagagtgggtggggtatctctgggcoogtgaggaggtcactgtcgttca
1	ENSE00000868271	153,724,868	153,726,166	-	0	1,299	GCCAGGCTGCGGAGCGGACGGACGCGCCTGGTGCCCGGGGAGGGGCGGCCACCGGGGAG GAGGAGGAGGAGAAGGTGGAGAGGAGAGACGCCCTCTGCCCGAGACCTCTCAAGGCC CTGACCTCAGGGGCCAGGGCACTGACAGGACAGGAGAGCCAGTTCTCTCACTTGGGCTG CCCGAAGAGGCCGGAACCTGGAGGGCCCTGAGGCCACCGCACCAGGGGCCAGCAACA CCCCGGGGCCCTAAGCGACAGTCTCAGGGGCCATCGCAAGGTTTCCAGTTGCCTAGACA ACAGGCCCGAGGTGAGAGCAACATCCTTCCAGCCACTGCTCCACTGCTGCCCGAGGC ACCAGGCCAGTCCCTACGCGGCGAGCCAGGCCAGGTGACATGCCGTGCTCTCCAGGCC CGGCCCTGCGGGGGAACAGCTGAAGCGCACGGCCGTGCTCTGGCCCTCGCGGCTAT GAGGCCACAAAGTCTACCCCTTGGTGGCCAGTGCCTGGCCCGGCGAGGGTCTTCAG GGCCCGCGGGGAGCCCAAGCAGGAGGCTCCGGGCTGGCGGGGCCAAAGCTGGCATG AACCGGGTATTCCTGAGCGGCTCCTGTGGCTCCTGGGCTGCTGTTCCCGCGGGTCTG TGCCGGGAGACGGGCTGCTGGCCCTGCACTCGGCCGCTTGGTGAGCGCAGCTTCTG TCGGTGTATGTGGCCCGCTGGACGGAAGGCTGGCCGCTGCATCGTCCGCAAGGACCG CGGGCTTTGGCTGGCAGCTGCTGCACTGGCTCCTCATGCCCTCCCTGCTACCTTCGT AACAGTGCCATCCGTTACCTGGAGGCGCACTGGCCCTGTGTTCCGAGCGCTCTGGTG GCCACGCTACCGCTCTACTTCTCCAGCAGACCTACTACGGGTGAGCAACATGGAC GGCGGGCTTCGCAACCTGACCACTCTCTGACGAGGAGCTGGTGGCTTTGGCGCTCT GTGGCCACCTCTACTCCAACTGACCAAGCCACTCCTGGAAGTGGCTGTGACTTCTTAC ACCTGCTTCGGGCGGCCGCTCCGCTGGAGCGGCGCACAGCTGGCCCTCGGCCATCGCC GGCTCGTGGTGTTCCTCAGGCGCAACGTGCTCGGGCTTCTCGCCCAAGTTCCGGGAG CTGGTGGCAGAGGAGCGCGCGGAGGAGGAGCTGGCTACATGCACTCGCTGTGGTG GCCAAGCTCGAGGAGATCGCCTTCTATGGGGCCATGAG
	Intron 1-2	153,726,167	153,729,231			3,065	gtggggcaggttgggtgcccggga.....tctctgtgtgtgtcacccccgag

🔗 Click on 'Sequence - cDNA' in the side menu.

The spliced transcript, as well as its coding sequence and translation are shown with sequence variants drawn. The page can be simplified with 'Configure this page'.

421 CGGCCCTGGCGGGGGAACACGCTGAAGCGCACGGCGTCTCTGGCCCTCGCGGCCTAT
 22 CGGCCCTGGCGGGGGAACACGCTGAAGCGCACGGCCGTCTCTGGCCCTCGCGGCCTAT
 8 -R--P--W--R--G--N--T--L--K--R--T--A--V--L--L--A--L--A--A--Y--
 481 GGAGCCACAAAGTCTACCCCTTGGTTCGCGCAGTGCCTGGCCCGGGCCAGGGGTCTTCAG
 82 GGAGCCACAAAGTCTACCCCTTGGTTCGCGCAGTGCCTGGCCCGGGCCAGGGGTCTTCAG
 28 -G--A--H--K--V--Y--P--L--V--R--Q--C--L--A--P--A--R--G--L--Q--

Click on 'External References - General identifiers' in the side menu.

On the 'General identifiers' page cross-references to other databases are shown that contain entries that correspond to the ENST00000218104 sequence.

General identifiers ⓘ

This transcript corresponds to the following database identifiers:

External database	Database identifier
CCDS	CCDS14728.1 [view all locations]
European Nucleotide Archive	BC015541 [align] [view all locations] BC025358 [align] [view all locations] U52111 [align] [view all locations] Z21876 [align] [view all locations] Z31006 [align] [view all locations] Z31007 [align] [view all locations] Z31008 [align] [view all locations] Z31009 [align] [view all locations] Z31010 [align] [view all locations] Z31348 [align] [view all locations]
HGNC transcript name	ABCD1-001 ATP-binding cassette, sub-family D (ALD), member 1 [view all locations]
Havana translation	OTTHUMP00000025960 [view all locations]
Human Protein Atlas	HPA035214 [view all locations] HPA035214 [view all locations]
INSDC protein ID	AAH15541.1 [align] [view all locations] AAH25358.1 [align] [view all locations] CAA79922.1 [align] [view all locations] CAA83230.1 [align] [view all locations]
RefSeq mRNA	NM_000033.3 [align] [view all locations]
RefSeq peptide	NP_000024.2 [align] [view all locations] (Target %id: 100; Query %id: 100) ATP-binding cassette sub-family D member 1 [view all locations]
UCSC Stable ID	uc004ff.2 [view all locations]
UniParc	UPI0000000DF5 [view all locations]
UniProtKB/Swiss-Prot	P33897 [align] ATP-binding cassette sub-family D member 1 [view all locations]

For example, ENST00000218104 matches the P33897 protein sequence in the UniProtKB/Swiss-Prot database (<http://www.uniprot.org/>) and the NP_000024.2 protein and NM_000033.3 mRNA sequence in the RefSeq database (<http://www.ncbi.nlm.nih.gov/refseq/>).

Click on 'Ontology - GO table' in the side menu.

Gene Ontology (GO) terms (<http://www.geneontology.org>) associate proteins to biological process, molecular function and cellular component terms.

GO table ⓘ

- [GO: Biological process](#)
- [GO: Cellular component](#)
- [GO: Molecular function](#)

Descendants of GO: Biological process

Accession	Term	Evidence	Annotation Source	GOSlim Accessions	GOSlim Terms
GO:0006200	ATP catabolic process	IDA		GO:0034641 GO:0034655 GO:0009056 GO:0008150 GO:0044281	cellular nitrogen compound metabolic process nucleobase-containing compound catabolic process catabolic process biological_process small molecule metabolic process
GO:0006635	fatty acid beta-oxidation	IDA		GO:0009056 GO:0008150 GO:0044281 GO:0006629	catabolic process biological_process small molecule metabolic process lipid metabolic process
GO:0006810	transport	IEA	InterPro:FA_transporter	GO:0008150	biological_process
GO:0007031	peroxisome organization	NAS		GO:0008150	biological_process

The 'biological process' terms indicate that the ABCD1 protein plays a role in fatty acid transport and catabolism. The 'cellular component' terms indicate the ABCD1 protein is located in the peroxisomal membrane.

Let's have a look at the region on the chromosome, and neighbouring genes.

🖱️ Click on the 'Location' tab.

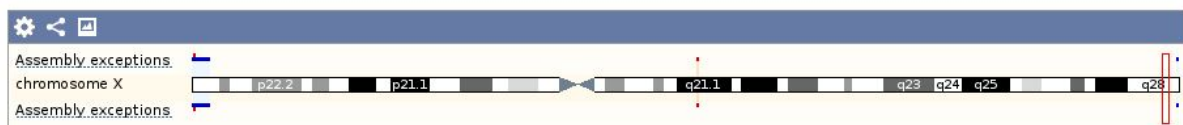
This leads us to the 'Region in detail' page under the 'Location' tab.

The Location tab

The 'Region in detail' page shows the genomic neighbourhood of the *ABCD1* gene. It consists of three parts.

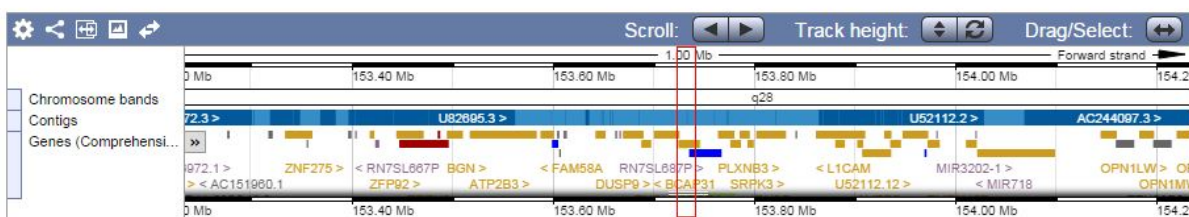
First, the complete chromosome.

Chromosome X: 153,724,868-153,744,762



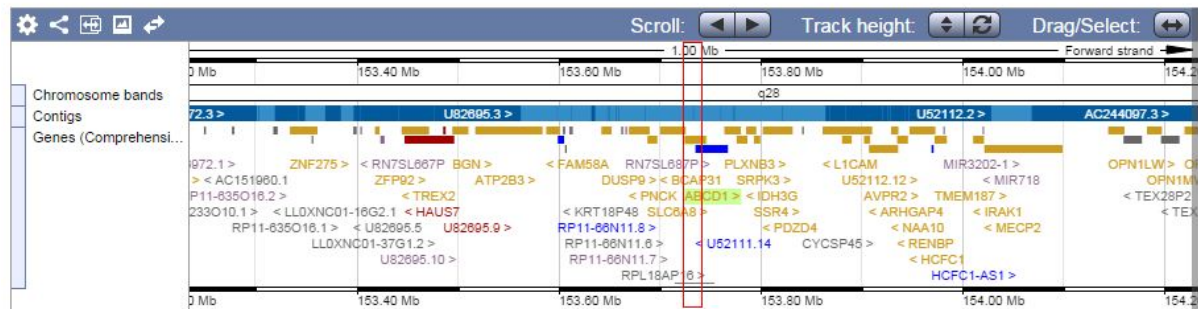
Second, the 1 Mb region around the gene of interest.

Region in detail ⓘ



Drag down the bottom to reveal *ABCD1*.

Region in detail



This display is scrollable. Either use the 'Scroll' arrows click and drag the image in the same way as Google Maps. Zoom in by clicking the 'Drag/Select' icon, selecting the region of interest with your mouse and subsequently clicking 'Jump to region' in the resulting pop-up.

Third, the region of interest. In our case this is the *ABCD1* gene.



By default, the data tracks drawn are:

- 38 way GERP elements (the 'constrained elements', which are regions of high conservation based on comparison of sequence across 38 species)
- Human cDNAs (cDNA sequences aligned to the genome)
- CCDS set (transcripts in the Consensus Coding Sequence Set)
- Genes (GENCODE)

- Contigs (the genome)
- 1KG-All-common (Variants from the 1000 Genomes project with population frequency >1%)
- All phenotype-associated variants
- 1000 Genomes High Quality Structural Variants
- MultiCell regulatory features (sequences that may be involved in gene regulation)
- %GC (reflects GC content vs AT)

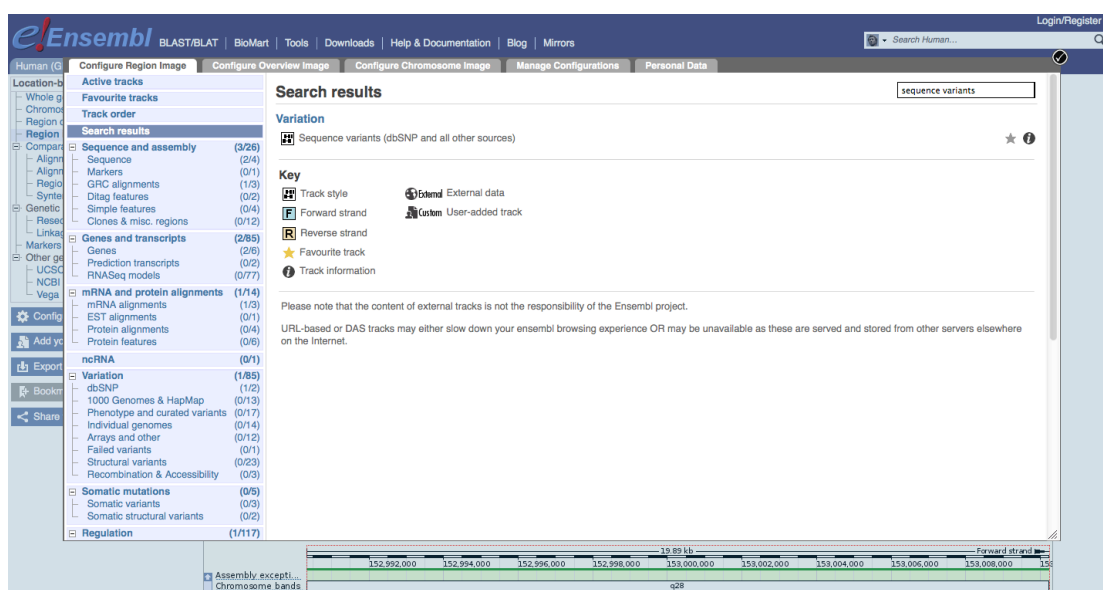
There are several ways to navigate this display:

- zoom in and out by using the [+/-] slider
- zoom in by drawing a box around the region of interest and subsequently clicking 'Jump to region' in the resulting pop-up
- moving up- and downstream with the single and double arrows next to the [+/-] slider.
- going to a particular region by changing the coordinates in the 'Location' text box or by searching for a gene using the 'Gene' text box (which has auto completion)

Datasets (or tracks) can be added to the display using [Configure this page]. On the configuration page all available tracks are grouped in the left-hand menu. It is also possible to search for tracks using the 'Find a track' text box.

For example, to add protein alignments from UniProt to the display:

- ✎ Click [Configure this page] in the side menu.
- ✎ Type 'UniProt' in the 'Find a track' text box.
- ✎ Select 'Proteins (mammal) from UniProt'. Choose 'Normal'.
- ✎ Click (✓).



A new track, 'UniProt (mammals)', has now been added to the display.



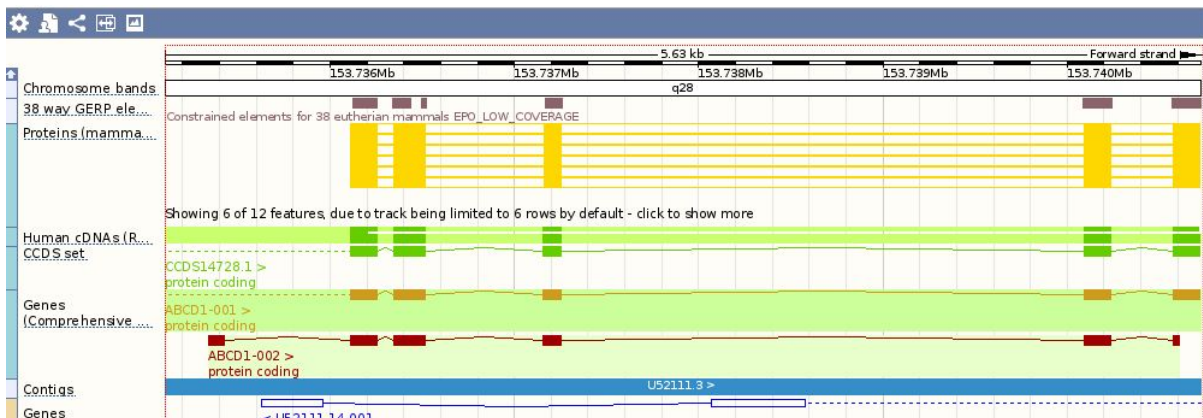
To turn the added track off again:

- ☞ Hover over the track name.
- ☞ Click on the 'Turn track off' icon (x) in the pop-up.

Tracks can be moved by clicking on the bar in front of the track name and dragging the track to the desired location.

To zoom in, you can click and drag your mouse around a region.

- ☞ Zoom in to ABCD1-002



At the top of the display (circled in the image above) several icons are shown, some of which can also be found on other displays:

- Configure this image: add/delete tracks (same as [Configure this page] button in the side menu).
- Manage your custom tracks: add your own data (same as [Add your data] button in the side menu)
- Share this image: create a URL that can be shared with others without the need to tell them how to configure the page
- Resize this image: resize the image
- Export this image: export the image in various formats (PDF, PNG etc.)

BLAST/BLAT

Finally, we will do a sequence search. For this we need some sequence. To this end, we will take the sequence of the region we are looking at.

- ☞ Click [Export data] in the side menu.
- ☞ Click [Next>].
- ☞ Click on 'Text'.

This gives us sequence in FASTA format.

```
>X dna:chromosome chromosome:GRCh38:X:153735103:153740728:1
TCCTTTTCATGGCTGAATAATATTCATTGTGTGCATAGACCACAATTTCTTTATCCATT
CATCCCTTGATGGACATTTTGGGTTTCTTCATGTTTTGGCTATTGTGAATAACACTGCTG
TGAACATCCATGGACAAGTCTCTATGTGTGCAGATATTTTCGTTTCTCCTGGGTGTGTAG
CTAGGAGTAGAATTGCCAGGTCACATGGTAACTGGACGTTTCACTTTTGAGGAGCTGCG
AGACTGTTCTCCACAGTGGCTGCCCATTTTACCTTCCCGCCAGCAGTGTGGAGGGTTC
CACCTTTTCATCGTGGCTAGCACTGGTTATCATCTCCTTTGTATTCTAGCCACCTAGTGG
GTGTGAGGCAGTATCTCTTGGTGGTTTTGATTTGCATTTCCCTGATGACTAATGACGCTG
AGCCTCTTTTGATGTGTTGAGTGGCCATTTGTATGTCTTTTGGAGAAATGTCTGTTCA
CGTCCTTCGCCCATGTGTGATCGGGTTATCTCTGTCGCTGAGTTGTAAAAGCTCTTTGTA
TATTCTGGATACTGCACCTCATCAGATGTGTGGTTACCAGTCCAGAGTTTTCTCCAG
```

- ☞ Select and copy the sequence.
- ☞ Go back to the browser.
- ☞ Click on the 'BLAST/BLAT' link on the toolbar at the top of the page.
- ☞ Paste the sequence in the 'sequence data' text box.
- ☞ Select 'Homo_sapiens' as the species to search against.
- ☞ Select 'BLAT' as the search tool.

We use BLAT because we expect an exact match (as a rule of thumb use BLAT when you are searching against the same species your query sequence is from and use BLAST when you are searching against another species).

- ☞ Click [RUN>].
- ☞ Click [View results].

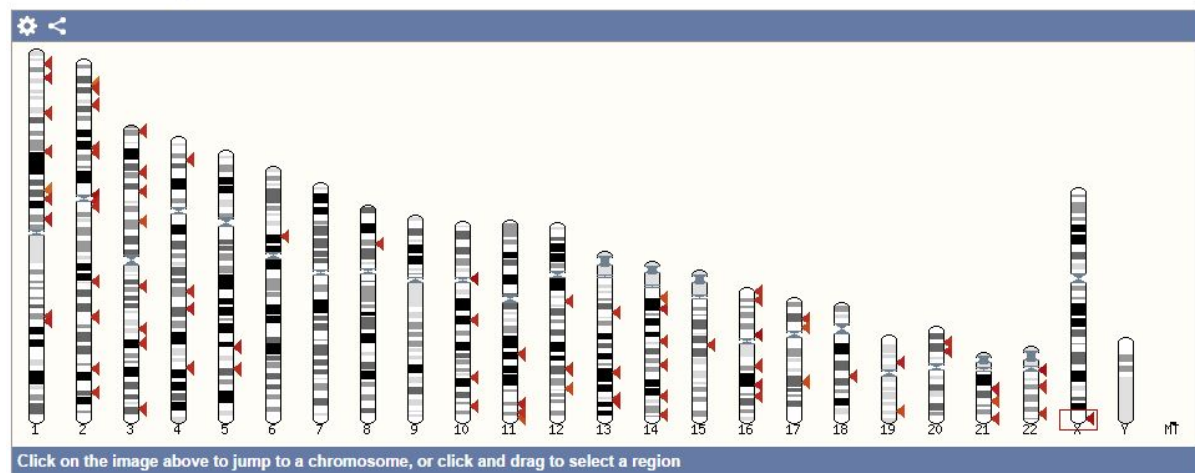
Results table:

Show	All ▾	entries	Show/hide columns							Filter	
Genomic Location	Orientation	Query name	Query start	Query end	Query ori	Length	Score ▾	E-val	%ID		
X:153735103-153740728 [Sequence]	Forward	X [Sequence]	1	5626	Forward	5626	11096.0	0.0e+00	100.00 [Alignment]		
X:154281643-154282028 [Sequence]	Forward	X [Sequence]	3195	3580	Forward	386	699.0	1.0e-201	95.08 [Alignment]		
X:154244522-154244907 [Sequence]	Forward	X [Sequence]	3195	3580	Forward	386	699.0	1.0e-201	95.08 [Alignment]		
X:154206713-154207098 [Sequence]	Forward	X [Sequence]	3195	3580	Forward	386	699.0	1.0e-201	95.08 [Alignment]		
X:154169585-154169970 [Sequence]	Forward	X [Sequence]	3195	3580	Forward	386	699.0	1.0e-201	95.08 [Alignment]		
16:32478096-32478329 [Sequence]	Reverse	X [Sequence]	5393	5626	Forward	234	443.0	1.2e-124	96.58 [Alignment]		
KI270728.1:624428-624661 [Sequence]	Forward	X [Sequence]	5393	5626	Forward	234	440.0	8.0e-124	96.15 [Alignment]		

This shows all hits: the length of the match, score, e-value and %ID (% identity).

Scroll down past the results table to find the BLAST matches drawn graphically on chromosomes. The best match has a box around it.

HSP distribution on genome: 



Look at the table again. The first hit is the best one in the table (the length matches the full query sequence, the e-value is 0, and the identity is 100%).

Genomic Location	Orientation	Query name	Query start	Query end	Query ori	Length	Score	E-val	%ID
X:153735103-153740726 [Sequence]	Forward	X [Sequence]	1	5626	Forward	5626	11096.0	0.0e+00	100.00
X:154281643-154282028 [Sequence]	Forward	X [Sequence]	3195	3580	Forward	386	699.0	1.0e-201	95.08
X:154244522-154244907 [Sequence]	Forward	X [Sequence]	3195	3580	Forward	386	699.0	1.0e-201	95.08

Click on the location of the match (circled above).

This leads us back to the 'Region in detail' page, to which now a new track named 'BLAT/BLAST hits' has been added, which shows the best hit in red.

