# **Tutorial: Ensembl walkthrough**

In this tutorial you will walk through Ensembl using an example. You 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 <u>http://en.wikipedia.org/wiki/Adrenoleukodystrophy</u>).

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



## Searching

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

 ${}^{\prime \! \oplus}$  Select 'Search: Human' and type 'abcd1' in the 'for' text box.

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



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 (<u>http://www.gencodegenes.org/</u>).. 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.

									Login/Registe
CEnsembl BL	AST/BLAT   BioN	1art   Tools   Dowr	nloads   H	elp & Documen	tation   More 🔻	9	- Search Huma	an	(
Human (GRCh38) V Location: >	(:153,724,868-153	,744,762 Gene: A	BCD1						
Gene-based displays	Gene: Al	BCD1 ENSG0000	0101986						
- Transcript comparison	Description		ATP	-binding casset	e, sub-family D (A	ALD), member	1 [Source:HGN	IC Symbol;Acc:HGN	C:61]
- Supporting evidence	Synonyms		adre	noleukodystrop	hy, ALD, ALDP, A	MN			
Secondary Structure	Location	Location Chromosome X: 153,724,868-153,744,762 forward strand.							
- External references	INSDC coord	INSDC coordinates chromosome:GRCh38:CM000685.2:153724868:153744762:1							
- Expression - Comparative Genomics - Comparative June Genomics - Comparative Genomics - C	Transcripts		This	gene has 3 trar	nscripts (splice va	riants) Hide t	ranscript table		
Gene tree (image)     Gene tree (text)	Show/hide	columns (1 hidden)					Filte	r.	
- Gene tree (alignment)	Name 🍦	Transcript ID 👙	Length 👙	Protein 🝦	Biotype 🝦	CCDS 🝦	RefSeq 👙	Flags	\$
<ul> <li>Gene gain/loss tree</li> <li>Orthologues (55)</li> </ul>	ABCD1-001	ENST0000218104	3664 bp	745 aa ( <u>view</u> )	Protein coding	CCDS14728	<u>NM_000033</u> NP_000024	GENCODE bas	sic
Protein families (3)	ABCD1-003	ENST00000370129	1016 bp	227 aa ( <u>view</u> )	Protein coding	14	-	GENCODE bas	sic
<ul> <li>Phenotype</li> <li>Genetic Variation</li> </ul>	ABCD1-002	ENST00000443684	668 bp	223 aa ( <u>view</u> )	Protein coding	8		CDS 5' and 3' incor	mplete

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

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

	153.72Mb	153.73Mb	153.74Mb					
Genes (Comprehensive		ABCD1-001 >	(I					
		ABCD1-003 >	ABCD1-002 >					
Contigs	U52	111.3 >	U52111					
Genes (Comprehensive	BCAP31-001 protein coding		□□ < U52111.14-001 antisense					
	< BCAP31-002 protein coding							

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						
<ul> <li>There are three transcripts, all protein coding</li> <li>ABCD1-001 is gold, a symbol of high quality</li> </ul>						

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.

# Marked-up sequence 0



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.

A 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

Let's find out more about ABCD1.

A 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.

					Filter	
Transcript ID	CCDS	UniProtKB/ Swiss- Prot	RefSeq peptide	RefSeq mRNA	Vega transcript	UniProtKB/ TrEMBL
ENST0000021810	4 <u>CCDS14728.1</u>	P33897	NP_000024.2	<u>NM_000033.3</u>	OTTHUMT0000061041	
ENST000037012	2				OTTHUMT0000061043	A6NEP8
ENST0000044368	4				OTTHUMT0000061042	

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

A 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 0

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
CADDS	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)
<mark>4</mark> 75	Show	ALL variations with a phenotype annotation			*
1	Show	ADRENOLEUKODYSTROPHY	View on Karyotype	-	OMIM
1	Show	ADRENOMYELONEUROPATHY	View on Karyotype	1.75)	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

A Click on 'GO: Biological process' in the side menu.

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

SAPOP: ARCO1 INSCOM								
e offer Place of Chapter	00101906							
Description	ATTP-binding cassette, sub-family D (ALD), member 1 ISource HGNC Svm	Not Acc HONC 61 67						
Synonyma	ALD. AMN, ALDP, ABC42, adrensleukodystrophy	AMI ALDP ABCD, scheenieukedustrachy						
Location	Chramosome X, 153 724 868 153 744 752 forward strand.	doctore X, 153 728,050 153 744,752 forward strand.						
	GRICK18 CM000685.2							
About this gene	This gave has 3 transcripts (splice variants), 56 orthologues, 3 pacalogues	, is a monitor of <u>1 Enzoni</u>	(protein family and its associated with)	5 phenotypes				
Transcripts	Hide transcript table							
art. Steachide columns			Pilot N	1				
Name A Transcript ID	he Protein 1 Translation ID 2 Biotome	DelPost 1 RelSec 1	flag					
ABC01-001 ENET2002021810	3664 745au ENSPIO202210104 Protein coding CCD211472047	P3393749 184 00013345	TEL1 GENCODE Basic APPRILIP	1				
ABCD1-003 ENST1000037012	1016 227au ENSP00200359147 Protein coding	AREPIG	TSL:2 GENCODE basic					
ABCD1-002 ENST2002044368	653 No protein - Processed transcript -		(TSL2)					
GO: Biological proce	55							
Chrone All (a) entries		Showhi	le columna					
Accession	Term		Exidence	Association Source	Transcript IDs	1142		
GO.00266354P	fatty acid beta-midation		IDA, KA	UnProblems Swiss Prot P33007-#, UnProblems Prot P31007-#	ENST8000370129 ENST8000218154	<ul> <li>Baarch I</li> <li>View on</li> </ul>		
CO.MINELOF	transport		EA	httmPtp:ABC1_TM_dom@	ENST00101370129 ENST00101218104	Search     Search		
an manage			The Mark	IL D. MITCH. D. COMPT. O. L. D. MITCH. D. C. DOMATO				
00.0010010	peroxitome organization		No. 164	CHEMICED WHEFTER FORM P. CHEMICAL PROPERTY OF FORE	EN51000123E104	<ul> <li>Search L</li> <li>View an.</li> </ul>		
60.0015910/P	peroxinome organization peroxinomal long-chain fatty acid import		IGLIEA	UnProbability of the State of the State of the State of the State of State	ENGIDEERIS	<ul> <li>Starch I</li> <li>View an</li> <li>Starch I</li> <li>Wew an</li> </ul>		
60.001911/0 60.001911/0 60.001911/0	perconsorma organization perconsormal itempichalin fatty acidi import perconsormal membrane transport		IGLIEA NAS	Construction and a set of the set	Englisson23mm	- Skarchi - View m. - Bisarchi - Waw mi - Starchi - Starchi		
60.0015120* 60.0015120* 60.0015120*	personanan organisation personanal long-chain fatty acid import personanal membrane transport fatty acid beta-osidation using acyl-CoA andese		GLIEA NAS TAS	Und State Some First 2000 (* Und State State State First 2000) (* Und State State State State State (* Und State State State State State (* Und State	Existingioni 218/184 Existingioni 218/184 Existingioni 218/184 Existingioni 218/184	<ul> <li>Search J</li> <li>Wew m.</li> <li>Bearch J</li> <li>Wew m.</li> <li>Bearch B</li> <li>Wew m.</li> <li>Search B</li> <li>Wew m.</li> <li>Search B</li> </ul>		
CO.001010 CO.001010 CO.001010 CO.001000 CO.001000 CO.001000	perconstrue lagerization perconstruent ling chain faith (in year) perconstruent membrane transport faithy acid bete-excludion using acyl CoA windles unsaturated faithy acid metabolic process		IGLIEA NAS TAS TAS	United Systems that Court of United Systems to The Court of United Systems That Court of United Systems and the United Systems That COURT of United Systems that COURT of United States that COURT of	ENGINARIZINA ENGINARIZINA ENGINARIZINA ENGINARIZINA ENGINARIZINA ENGINARIZINA	<ul> <li>States 1)</li> <li>View an</li> <li>Barech 1</li> <li>View an</li> <li>Barech 2</li> <li>View an</li> <li>States 1</li> <li>View an</li> <li>States 1</li> <li>View an</li> <li>States 3</li> </ul>		
CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR	pensanne genezioni pensanne lang-bala (bry acid inport pensanne mediante insegnat fatty acid bala-anditaria using acyl Cub solates unastrated fatty acid mediatic process algha invitere: acid mediatic process		101.16A 103.16A 1045 1045 1045	ロビロクロロロビコルになっています。 はこれは2010年1月1日、2011年、日本の日本の日本の日本の日本の日本 はこれは2011年、日本の日本の日本の日本の日本の日本の日本の日本の日本の日本 についていたいたいたいであっていたいであっいいいいであっいいいいいのであっいいいいいいのであっいいいいいいいいいい	Ensistementariana Ensistementariana Ensistementariana Ensistementariana Ensistementariana Ensistementariana	Stateful I     Wow and     Stateful I     Wow and     Stateful I     Work and     Stateful I     Work and     Stateful I     Stateful I     Work and     Stateful I     Stateful     Stateful I     Stateful     Statefu		
0.0001200 0.00111100 0.00111100 0.0001000 0.0001000 0.0001000 0.0001000	versional of galaxies for y set of equit previous of environment transport (bitly set if the environment using any LCA institute unanament (bigg set in environment) and plan bitmices in environment trage-share (bitly and catalostic process trage-share (bitly and catalostic process		101 IEA 1145 1145 1145 1145 1145 1145 1145 114	Alexandrovani (Colome Viendoria) analogo (Colome Viendoria) Viendoria Science Intel 2001 P Mahard Science Intel 2001 P	Ensistemerziste Ensistemerziste Ensistemerziste Ensistemerziste Ensistemerziste Ensistemerziste Ensistemerziste	Barach     Who am     Barach     Who am     Barach     Statesh     States		
0.0000000 0.000000 0.000000 0.000000 0.000000	versional in particulars previous data good a flag y set of import previous and indexident strengent fetty and table-indexide uning any GCA indexide unsubsets flag yood invalidate process in rigo should be indexide process incego should be indexide process incego should be indexide process		63 MBM 62, 15A 146 146 146 146 146 146 146 146 146 146	Australization County Viendon Australia Markalidaen Int 2000 P United Science Int 2000 P		المعند - (ماندین - (ما دین - (ما دین - (ما دین - (ما دین - (ما دی		
0.0000000 0.000000 0.000000 0.000000 0.000000	venuemen ingenizionen personamente lange obes deng van it mopen personamente more tenengen detty moti heter-workdere eninge moje/CoA-molene unsecuented dang soch motiokle process digele insolver, can it motiokle process ingenizionen detty work candide process energisches detty work candide process insolute, cand motiokle process		50 MM 52 KA 946 946 949 949 949 949 949 949 949 949	Australization County Viendon Annual Park 2000 V Annual Park 2000 V Annual Park 2000 V Annual Park 2000 V	Charlanner, 2014 Charlanner, 2014 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015 Charlanner, 2015	<ul> <li>Sach</li> <li< td=""></li<></ul>		
0.00000109 0.0000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009	versional of particulars presentiated in endoting language of particulars (endoting and the endoting language) (endoting and endoting process and endoting and endoting process (endoting and endoting process		50 MM 52 KA 949 949 949 949 949 949 949 949 949 94	Lacing and an		Lanza - Lanza - Lan		
0.0000109 0.0000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009 0.000009	vennenna fepatoson prevensión fejado sefar yas el reput presidente fejado sefar yas el reput delay sol des exectés reagent delay sol dela condeció proces dela poste dela post dela poste la regulada dela post dela poste en y largo dela dela poste condeci post matidade proces condeci post matidade postes condeci poste dela postes		43, 163 44, 164 144 144 144 144 144 144 144	Lacing Control (Lacing) & Version (Lacing) & Versio		<ul> <li>Statushi - (San San San San San San San San San San</li></ul>		

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.

The 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.

E	Download sequence	💉 BLAST	this sequence	2				
Ke	ey .							
1	-							
Ð	kons/Introns	g sequence	Intron seq	uence		ranslated	sequence	
	UIR		]					
\$	Show All 🔻 entries				5	Show/hide	columns	Filter
N	o. Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence	a
	5' upstream sequence							ggaagagtgggtggggattotgggooggtggaggagtoactgtogottoa
1	ENSE00000868271	153,724,868	153,726,166	-	0	1,299		SCCAGGCTGCGGAGCGGACGGACGCGCCTGGTGCCCGGGGAGGGGCGACCGGGGGA SAGGAGGAGAAAGGTGAGAGGAAGAGCGCCCCCTCTGCCCGAGACCTCTCAGGGC CTGACCTCAGGGGCCAGGCACAGGCACGGCCGCCCCCCCGGAGACCTACTGGGCT CCCGGAGGGCCAGAGCACATCTCAGGGGCCACGCCACGGGCCCAGGGCCCAGCAC ACCGCCCGGGGCCAGAGCACATCTCCAGGGCCACGCCGCCCCGGCCCAGGCCCAGCCA ACCGCCCCAGTCCCTAGCGGCAGCCAGGCCAGGCCGCCTCCGGCCCCAGGCC ACCAGCCCCAGTCCCTAGCGGCAGCCAGGCCGGCCTCCGGCCCAGGCCCAGGCC CCCGCGGGGAACACACTCTCCAGCGCCAGGCCGGCCCGGCCCAGGCCCAGGCC ACCAGCCCCAGGGCAGACACACCCTGCCCCGGGCCGGCCCGGCCCGGGCCAGGCC CCCGCGCGGGGAACACCCTCGAGCGCCGCGCCGCGCCGGCCG
	Intron 1-2	153,726,167	153,729,231			3,065		gtggggcaggttggggtgccgggcatototgtgtotgtcaccocccgcag

Exons 0

A 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 0

This transcript corresponds to the following database identifiers:

Show All 🔻 entries		Filter	
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	OTTHUMP0000025960 [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 [Target %id: 100; Query %id: 100] [align] ATP-binding cassette sub-family D member 1 [view all locations]		
UCSC Stable ID	uc004fif.2 [view all locations]		
UniParc	UPI000000DF5 [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 (<u>http://www.uniprot.org/</u>) and the NP\_000024.2 protein and NM\_000033.3 mRNA sequence in the RefSeq database (<u>http://www.ncbi.nlm.nih.gov/refseq/</u>).

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

\$ < ⊡					
Assembly exceptions	<u>+</u> 0				Π.
chromosome X Assembly exceptions	p22.2	<u>12181</u>	<u>q21.1</u>	q23 q24 q25	q28
Assertibly exceptions	10-				.Ц.

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

#### Region in detail 0

🌣 < 🖻 🖬 🏕			Scroll:	Track height	nt: 🚺 😥 Dra	ag/Select: 😝
				2		Forward strand
	р Мь	153.40 Mb	153.60 Mb	153.80 Mb	154.00 Mb	154.2
Chromosome bands				q28		
Contigs	72.3>	U82695.3 >			U52112.2 >	AC244097.3 >
Genes (Comprehensi	» I I I					
	972.1 > ZNF275 > > < AC151960.1	< RN7SL667P BGN > ZFP92 > ATP2B3 >	< FAM58A RN7SL687P > DUSP9 > < BCAP3	PLXNB3 > < L1CAM 1 SRPK3 > U52112.1	MIR3202-1 > 2 > < MIR718	OPN1LW > O OPN1M
	р МЬ	153.40 Mb	153.60 Mb	153.80 Mb	154.00 Mb	154.2

Drag down the bottom to reveal ABCD1.

### Region in detail ()

🌣 < 🕀 🖬 🏕			Scroll:	Track heig	ght: 🚺 🔁 Dra	ag/Select: 😝
			1.00	Mb		Forward strand -
	3 Mb	153.40 Mb	153.60 Mb	153.80 Mb	154.00 Mb	154
Chromosome bands				q28		
Contigs	72.3>	U82695.3 >			U52112.2 >	AC244097.3 >
Genes (Comprehensi	1 1 1					
	1972.1 > ZNF27 > < AC151960.1 P11-635016.2 > 233010.1 > <ll0xnc RP11-635016.1 LLC</ll0xnc 	5> < RN7SL867P BGN > ZFP92 > ATP283 < TREX2 01-16G2.1 < HAUS7 > < U82895.5 U82695.9 > XNC01-37G1.2 > U82895.10 >	< FAM58A RN7SL687P > DUSP9 > < BCAI < PNCK AB < RP11-66N11.8 > RP11-66N11.8 > RP11-66N11.8 > RP11-86N11.8 > RP11-86N11.7 > RP118AP163	> PLXNB3 > <l1cam P31 SRPK3 &gt; U52112 CD1 &gt; <idh3g avf<br="">&gt; SR4 &gt; <a <pdzd4 U52111.14 CYCSP45 &gt;</pdzd4 </a </idh3g></l1cam 	MIR3202-1> 12> < MIR718 R2> TMEM187> RHGAP4 < IRAK1 < NAA10 < MECP2 < RED8P < HCFC1 HCFC1-AS1>	OPN1LW> ( OFN11 < TEX28P2 < TE
-	0 Mb	153.40 Mb	153.60 Mb	153.80 Mb	154.00 Mb	154

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.

	152 725 000	152 727 500	152 720 000	152 722 500	.9.89 kb	152 727 500	152 740 000	
romosome bands	155,725,000	153,727,500	155,750,000	153,752,500	q28	155,757,500	153,740,000	133,742,500
way GERP ele			L 500 LOW 00	UEDA OF				
iman cDNAs (R	Constrained element	s for 38 eutherian ma	immais EPO_LOW_CO	VERAGE				
DS set			_			·		
	CCDS14728.1 >							
	procent county							
omprehensive	ABCD1-001 >							
	procent county					·		
	ABCD1-003	>			ABCD1-002 >			
otiac	protein cod	U52111.3	>		protein coding	U52111	3 >	
nugs								
omprehensive					< U52111.1	14-001		
G - All - common	A CONTRACTOR	16 41	a - 11510		antisense			
phenotype-as								
00 Genomes	No 1000 Genomes - High quality (structural variants) data for this region							
ultiCall								
								25 23

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.
- A Type 'UniProt in the 'Find a track' text box.
- A Select 'Proteins (mammal) from UniProt'. Choose 'Normal'.
- ∽ Click (✓).



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

Proteins (mamma						
Human cDNAs(R	Showing 6 of 12 features, due to track being limited to 6 rows	by default - click to show more				
CCD S set	CCDS14728.1 > protein coding					
Genes (Comprehensive	ABCD1-001 > protein coding					
	ABCD1-003 > protein coding	ABCD1-002 > protein coding				
Contigs	U52111.3 >		U52111.3 >			

To turn the added track off again:

Hover over the track name.

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.

 $^{\circ}$  Zoom in to ABCD1-002

		Forward strand 🗩						
	153.736Mb	153.737Mb	153.738Mb	153.739Mb	153.740Mb			
Chromosome bands	q28							
38 way GERP ele	Constrained elements for 38 eutherian mammals E	PO LOW COVERAGE						
Proteins (mamma								
	Showing 6 of 12 features, due to track being li	mited to 6 rows by default - (	click to show more					
Human cDNAs (R								
CCDS set								
	protein coding							
Genes								
(Comprehensive	ABCD1-001 >							
				- <u> </u>				
	ABCD1-002 >							
	protein coding		U52111 3 >					
Jontigs								
Genes	< U52111.14-001							

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.)