

These demonstrations will lead you through the main databases and tools for genetic variation at the NCBI.

**dbVAR**

There are phase II drug metabolism enzyme genes that are missing in some individuals, notably some glutathione S-transferase genes (GSTM1, GSTM2 and GSTT2) and a UDP glucuronosyl transferase (UGT2B17). We can identify variant calls from the 1000 Genomes project data in dbVAR to see how prevalent these deletions are. Go to the variation portal ([www.ncbi.nlm.nih.gov/variation/](http://www.ncbi.nlm.nih.gov/variation/)) page to access the dbVAR homepage.

Click the link to the Study Browser

Use the Filters on the right-hand side of the browser to filter for sequencing studies of human with more than 10,000 variant regions.

**dbVar: Study Browser**

Date	Publication	Variant Region Count	Variant Call Count
2015/03	Besenbacher et al. 2014	40,141	40,141
2015/02	Thareja et al. 2015	11,116	11,116
2014/10	1000 Genomes Consortium Phase 3	<a href="#">estd214</a> 62,855	6,623,477
2014/02	Boomsma et al. 2014	<a href="#">estd215</a> 28,083	28,083
2013/11	Pang et al. 2013	<a href="#">estd209</a> 471,817	471,817
2013/01	Wong et al. 2013	<a href="#">estd201</a> 36,558	312,665
2010/10	1000 Genomes Project Consortium et al. 2010	<a href="#">estd59</a> 228,871	2,158,011
2009/09	McKernan et al. 2009	<a href="#">estd197</a> 232,775	232,775
2008/11	Bentley et al. 2008	<a href="#">estd194</a> 504,912	504,912

Retrieve study estd214, the 1,000 Genomes Consortium Phase 3 study.

**estd214**

**Organism:** [Human](#)  
**Study Type:** Control Set  
**Submitter:** [Laura Clark](#)  
**Submitter URL:** <http://www.1000genomes.org/>  
**Description:** This study contains the structural variants from the combined release set which contains more than 79 million variant sites and includes not just biallelic snps but also indels, deletions, complex short substitutions and other structural variant classes. It is based on data from 2504 unrelated individuals from 26 different populations around the world.  
**Project:** [PRJEB6930](#)

**Detailed Information:** [Download 62855 Variant Regions](#), [Download 6623477 Variant Calls](#), [Download Both, FTP](#)

**Links**

- [Variants in this study](#)
- [Open Human in Taxonomy Browser](#)
- [BioProjects](#)

Source: NCBI

Follow the link to retrieve the “Variants in this study” and add GSTM1 to the search.

dbVar **estd214 AND GSTM1**  
[Save search](#) [Advanced](#)

**Display Settings:**  Tabular View [Send to:](#)

**Results: 3**

**Number of Variants: 3**

Variant Region ID	Type	Number of Variant Calls	Study ID	Organism	Clinical Assertion	Location	Genes in region
<a href="#">esv3587154</a>	copy number variation	2141	<a href="#">estd214</a>	human		<a href="#">GRCh37 (hg19) chr1: 110,224,019-110,246,280</a> , <a href="#">GRCh38 (hg38) chr1: 109,681,397-109,703,658</a>	GSTM1, GSTM2
<a href="#">esv3587155</a>	copy number variation	2156	<a href="#">estd214</a>	human		<a href="#">GRCh37 (hg19) chr1: 110,230,075-110,241,247</a> , <a href="#">GRCh38 (hg38) chr1: 109,687,453-109,698,625</a>	GSTM1
<a href="#">esv3587156</a>	copy number variation	23	<a href="#">estd214</a>	human		<a href="#">GRCh37 (hg19) chr1: 110,230,075-110,241,247</a> , <a href="#">GRCh38 (hg38) chr1: 109,687,453-109,698,625</a>	GSTM1

We have deletions and duplications here, some that cover both the GSTM1 and GSTM2 genes, which are in tandem on chromosome 1. Click on the link for the Variant Region ID for the one affecting both GST genes ([esv3587154](#)). In expanded form the dbVar browser show the variant region plus a large number of variant calls.

**esv3587154**

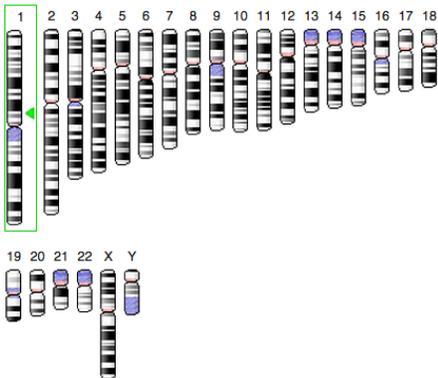
**Organism:** [Homo sapiens](#)

**Study:** [estd214 \(1000 Genomes Consortium Phase 3\)](#) **Variant Calls:** 2,141

**Variant Type:** copy number variation **Validation:** Not tested

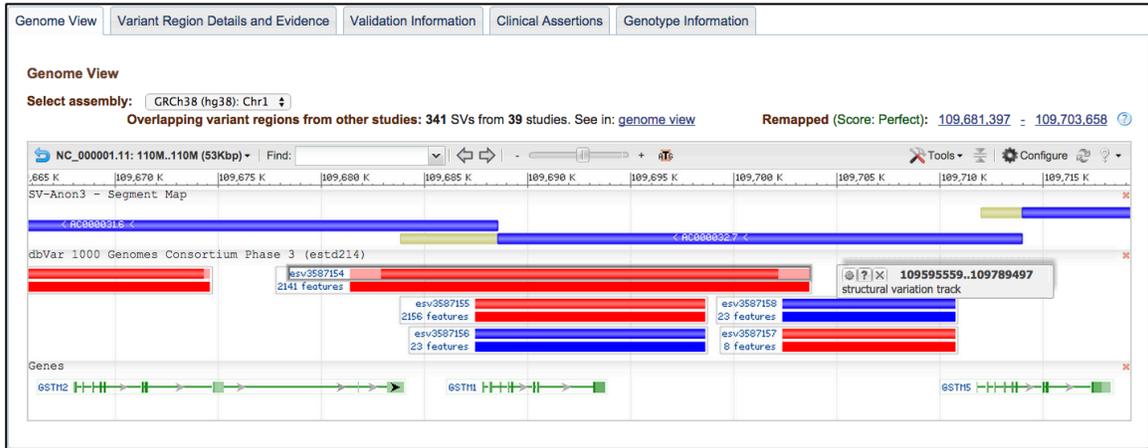
**Method Type:** Sequencing **Clinical Assertions:** No

**Submitted on:** GRCh37 (hg19) **Region Size:** 22,262



GRCh38 (hg38) [esv3587154](#)

You can use the Configure menu to set the display options for the structural variation track to “Show parent, Merge children”, making the display simpler. Use the zoom feature on the viewer to expand to include both GSTM1 and GSTM2 genes.



You can see the track legend by mousing over the head of the structural variation track and clicking on the question mark. The red graphic indicates that this is a deletion (copy number loss). Notice that the other common region, esv3587154, is also a copy number loss, but the less frequent esv3587156 is a copy number gain. You can click on the “Variant Region Details and Evidence” tab to see the individual level genotypes for the structural variation.

Variant Call ID	Type	Sample ID	Method	Analysis	Zygosity	Other Calls in this Sample and Study
essv10056817	copy number loss	<a href="#">SAMN00004622</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,029</a>
essv10056818	copy number loss	<a href="#">SAMN00004623</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,569</a>
essv10056819	copy number loss	<a href="#">SAMN00004625</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,494</a>
essv10056820	copy number loss	<a href="#">SAMN00004626</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,840</a>
essv10056821	copy number loss	<a href="#">SAMN00004627</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,576</a>
essv10056822	copy number loss	<a href="#">SAMN00004628</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,591</a>
essv10056823	copy number loss	<a href="#">SAMN00004631</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,605</a>
essv10056824	copy number loss	<a href="#">SAMN00004632</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,578</a>
essv10056825	copy number loss	<a href="#">SAMN00004633</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,544</a>
essv10056826	copy number loss	<a href="#">SAMN00004634</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,579</a>
essv10056827	copy number loss	<a href="#">SAMN00004635</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,666</a>
essv10056828	copy number loss	<a href="#">SAMN00004636</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,583</a>
essv10056829	copy number loss	<a href="#">SAMN00004637</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,499</a>
essv10056830	copy number loss	<a href="#">SAMN00004638</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,269</a>
essv10056831	copy number loss	<a href="#">SAMN00004639</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,601</a>
essv10056832	copy number loss	<a href="#">SAMN00004640</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,132</a>
essv10056833	copy number loss	<a href="#">SAMN00004641</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,811</a>
essv10056834	copy number loss	<a href="#">SAMN00004642</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,561</a>
essv10056835	copy number loss	<a href="#">SAMN00004643</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,449</a>
essv10056836	copy number loss	<a href="#">SAMN00006337</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,786</a>
essv10056837	copy number loss	<a href="#">SAMN00006338</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,630</a>
essv10056838	copy number loss	<a href="#">SAMN00004644</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,600</a>
essv10056839	copy number loss	<a href="#">SAMN00006339</a>	Sequencing	Read depth and paired-end mapping	Homozygous	<a href="#">2,641</a>
essv10056840	copy number loss	<a href="#">SAMN00006340</a>	Sequencing	Read depth and paired-end mapping	Heterozygous	<a href="#">2,632</a>

Click on the Sample ID for the first heterozygote in the table (SAMN00004626) to see details of the individual.

**Coriell HG00100**

**Identifiers** BioSample: SAMN00004626; SRA: SRS006841; Coriell: [HG00100](#); 1000G: HG00100

**Organism** [Homo sapiens \(human\)](#)  
 cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

**Attributes**

<b>population</b>	<a href="#">GBR</a>
<b>Population Description</b>	British From England and Scotland, UK
<b>Super Population Description</b>	European
<b>sex</b>	female
<b>culture collection</b>	<a href="#">Coriell:HG00100</a>
<b>family role</b>	unrelated
<b>Super Population Code</b>	EUR
<b>Coriell panel</b>	MGP00003
<b>DNA-ID</b>	HG00100

**Description** Human 1000 genomes individual HG00100

**Links** [DNA source](#)  
[dbSNP Batch ID 1061891](#)

**BioProjects** [PRJNA262923](#) 1000 Genomes Project phase3  
 Retrieve [all samples](#) from this project

[PRJNA60113](#) Exome sequencing of (GBR) British from England and Scotland HapMap population  
 Retrieve [all samples](#) from this project

[PRJNA41223](#) Whole genome sequencing of (GBR) British from England and Scotland HapMap population  
 Retrieve [all samples](#) from this project

There are only 100 variant calls in the table. You could get the complete set (VCF) from the 1000 genomes FTP directory. Notice that you can load all structural variants in the region into the dbVAR Genome Browser. This is an application that is very similar to the Variation Viewer that we will use later.

There are no clinical assertions associated with these 1000 Genomes data. You can search dbVar with another gene to find variants that have entries in ClinVar. Search for UGT2B17 from the main dbVar page. Filter your results by Object type "Variant" and Clinical Assertion "association".

Object Type  **Display Settings:**  Summary Send to:

Variant (1)

Study (0)

**Organism** [nsv513778](#)

human (1)

Bornean orangutan (0)

Sumatran orangutan (0)

**Variant** [nsid51](#)

copy number variation (1)

complex (0)

complex chromosomal mutation (0)

**Method Type**

Curated (1)

BAC aCGH (0)

Digital array (0)

**Clinical Assertion**

association (1)

**Filters activated:** Variant, association. [Clear all](#) to show 824 items.

**Variant type:** copy number variation

**Associated study:** [nsid51](#)

**Organism:** human

**Gene(s) in region:** [TMPRSS11B](#), [TMPRSS11E](#), [TMPRSS11F](#), [TMPRSS11BNL](#), [UGT2B15](#), [UGT2B17](#), [YTHDC1](#)

**Location information:**

Submitted: GRCh37 (hg19); 4: 68871643-69625838

Remapped: GRCh38 (hg38); 4: 68005925-68760120

**Validation status:** Not tested

**Clinical significance:** **association**

ID: 1272798 **variant**

**Related information**

ClinVar

Gene

OMIM

Pathways + GO

PubMed

Taxonomy

**Search details**

UGT2B17[Gene Name] AND ("VARIANT"[OBJ\_TYPE] AND "association"[clinical\_assertion])

**Recent activity**

Follow the link to ClinVar. This is a 150 Kb deletion record provided by OMIM that is associated with low bone mineral density.

**nsv513778**

**nsv513778**

Variant type: Deletion

Cytogenetic location: 4q13

Other names: 150-KB DEL

Links: dbVar: [nsv513778](#)  
OMIM: [601903.0001](#)

**Clinical significance**

**nsv513778** [Help](#)

Clinical significance: association

Review status: ★ ☆ ☆ ☆

Number of submission(s): 1

**Condition(s)**

Bone mineral density quantitative trait locus 12 [MedGen - OMIM]

[See supporting ClinVar records](#)

---

**1 Affected gene**

**UDP glucuronosyltransferase 2 family, polypeptide B17 (UGT2B17)** [Gene - OMIM]

🔍 Search ClinVar for variants within UGT2B17

🔍 Search ClinVar for variants including UGT2B17

**Assertion and evidence details** Go to:

[Clinical Assertions](#) [Evidence](#) [Help](#)

**Germline**

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter - Study name (Last submitted)	Submission accession
association (Dec 16, 2011)	classified by single submitter (literature only)	literature only	Bone mineral density quantitative trait locus 12 [MedGen   OMIM]	germline	<a href="#">PubMed (1)</a> <a href="#">[See all records that cite this PMID]</a>	OMIM (Dec 30, 2010)	SCV000028275

The above are association (possible linkage) results. A useful tool for accessing association results from the NHGRI GWAS catalog and dbGaP is the Phenotype Genotype Integrator (PheGenI).

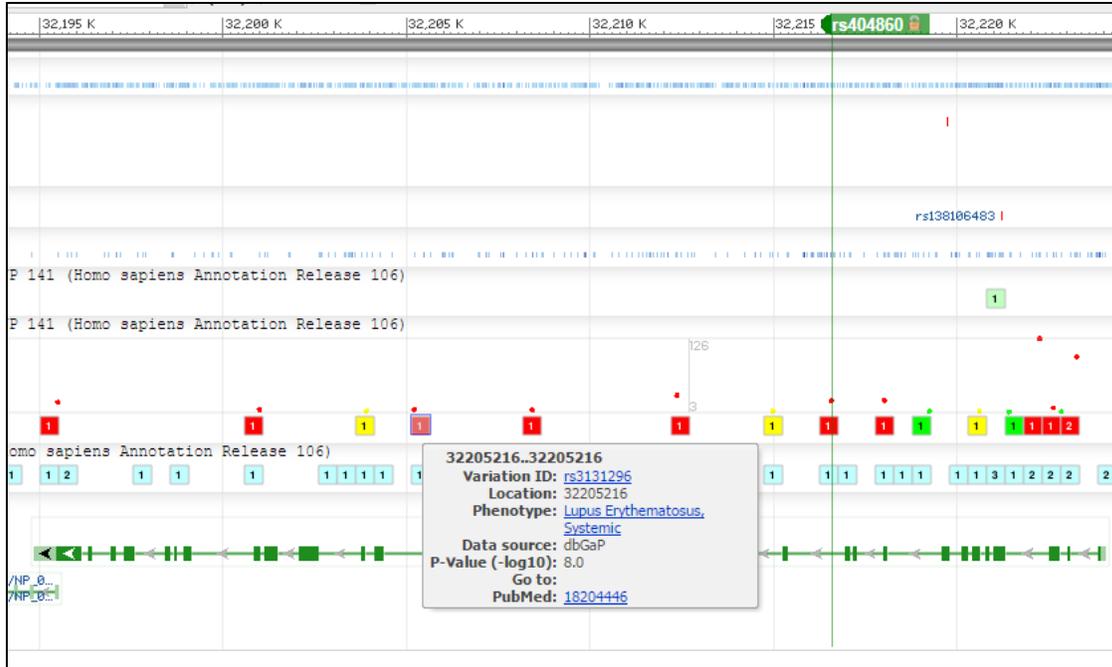
### Using PheGenI to find association results

Link to the Phenotype Genotype Integrator from the Variation Portal page. Search for asthma with the p value set to  $1 \times 10^{-8}$ . The results are in several sections. The most useful for our example is the Association Results.

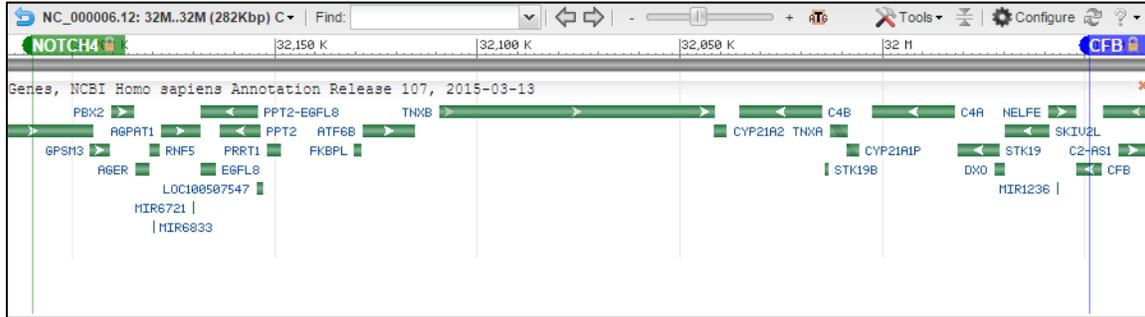
The Association Results shows the RefSNP identifier, the closest gene, the location on the chromosome, the p value and the source of the result (NHGRI or dbGaP) and any publications.

Association Results									
#	Trait	rs #	Context	Gene	Location	P-value	Source	Study	PubMed
1	Asthma	rs404860	intron	NOTCH4	6: 32,184,345	4.000 x 10 <sup>-23</sup>	NHGRI		21804548
2	Asthma	rs1837253	intergenic	SLC25A46, TSLP	5: 110,401,872	1.000 x 10 <sup>-16</sup>	NHGRI		21804548
3	Asthma	rs11078927	intron	GSDMB	17: 38,064,405	2.000 x 10 <sup>-16</sup>	NHGRI		21804549
4	Asthma	rs3771180	intron	IL1RL1	2: 102,953,617	2.000 x 10 <sup>-15</sup>	NHGRI		21804549
5	Asthma	rs204993	intron	PBX2	6: 32,155,581	2.000 x 10 <sup>-15</sup>	NHGRI		21804548
6	Asthma	rs10508372	intergenic	KRT8P16, TCEB1P3	10: 8,972,018	2.000 x 10 <sup>-15</sup>	NHGRI		21804548
7	Asthma	rs3129943	intron	C6orf10	6: 32,338,695	3.000 x 10 <sup>-15</sup>	NHGRI		21804548
8	Asthma	rs7775228	intergenic	HLA-DQB1, HLA-DQA2	6: 32,658,079	5.000 x 10 <sup>-15</sup>	NHGRI		21804548
9	Asthma	rs1837253	intergenic	SLC25A46, TSLP	5: 110,401,872	1.000 x 10 <sup>-14</sup>	NHGRI		21804549
10	Asthma	rs9273349	intergenic	HLA-DQA1, HLA-DQB1	6: 32,625,869	7.000 x 10 <sup>-14</sup>	NHGRI		20860503
11	Asthma	rs2786098	intron	CRB1	1: 197,325,908	2.000 x 10 <sup>-13</sup>	NHGRI		20032318
12	Asthma	rs1701704	intergenic	SUOX, IKZF4	12: 56,412,487	2.000 x 10 <sup>-13</sup>	NHGRI		21804548
13	Asthma	rs3129890	intergenic	HLA-DRA, HLA-DRB9	6: 32,414,273	5.000 x 10 <sup>-13</sup>	NHGRI		21804548
14	Asthma	rs3019885	intergenic	C8orf85, SLC30A8	8: 118,025,645	5.000 x 10 <sup>-13</sup>	NHGRI		21814517
15	Asthma	rs3019885	intron	SLC30A8	8: 118,025,645	5.000 x 10 <sup>-13</sup>	NHGRI		21814517
16	Asthma	rs946263	intergenic	CHI3L1, CHIT1	1: 203,165,381	9.738 x 10 <sup>-13</sup>	dbGaP	phs000123	11022011
17	Asthma	rs2153101	intergenic	CHI3L1, CHIT1	1: 203,168,474	9.738 x 10 <sup>-13</sup>	dbGaP	phs000123	11022011
18	Asthma	rs4950929	intergenic	CHI3L1, CHIT1	1: 203,160,126	1.321 x 10 <sup>-12</sup>	dbGaP	phs000123	11022011
19	Asthma	rs7686660	intergenic	FLJ44477, USP38	4: 144,003,159	2.000 x 10 <sup>-12</sup>	NHGRI		21804548
20	Asthma	rs2381416	intergenic	RANBP6, IL33	9: 6,193,455	2.000 x 10 <sup>-12</sup>	NHGRI		21804549
21	Asthma	rs3117098	intergenic	HNRNPA1P2, BTNL2	6: 32,358,513	5.000 x 10 <sup>-12</sup>	NHGRI		21804548
22	Asthma	rs9275698	intergenic	HLA-DQB1, HLA-DQA2	6: 32,687,973	5.000 x 10 <sup>-12</sup>	NHGRI		21804548
23	Asthma	rs7216389	intron	GSDMB	17: 38,069,949	9.000 x 10 <sup>-11</sup>	NHGRI		17611496
24	Asthma	rs2069408	intron	CDK2	12: 56,364,321	1.000 x 10 <sup>-10</sup>	NHGRI		21804548
25	Asthma	rs987870	intron	HLA-DPA1	6: 33,042,880	2.000 x 10 <sup>-10</sup>	NHGRI		21814517
26	Asthma	rs987870	nearGene-5	HLA-DPB1	6: 33,042,880	2.000 x 10 <sup>-10</sup>	NHGRI		21814517
27	Asthma	rs1342326	intergenic	RANBP6, IL33	9: 6,190,076	9.000 x 10 <sup>-10</sup>	NHGRI		20860503
28	Asthma	rs3771166	intron	IL18R1	2: 102,986,222	3.000 x 10 <sup>-9</sup>	NHGRI		20860503
29	Asthma	rs1101999	intron	PYHIN1	1: 158,932,555	4.000 x 10 <sup>-9</sup>	NHGRI		21804549
30	Asthma	rs9500927	intergenic	BRD2, HLA-DOA	6: 32,961,361	4.000 x 10 <sup>-9</sup>	NHGRI		21804548
31	Asthma	rs744910	intron	SMAD3	15: 67,446,785	4.000 x 10 <sup>-9</sup>	NHGRI		20860503
32	Asthma	rs3894194	missense	GSDMA	17: 38,121,993	5.000 x 10 <sup>-9</sup>	NHGRI		20860503
33	Asthma	rs2284033	intron	IL2RB	22: 37,534,034	1.000 x 10 <sup>-8</sup>	NHGRI		20860503

You can sort the table by the various column headers. Try sorting by position. Notice that many of the associated SNPs and nearby genes are in the human leukocyte antigen (HLA) region on chromosome 6, a region associated with immune and inflammatory response. The variant rs4048060 has the most significant p-value. Click through to the SNP record. There are a large number of variants in this gene with significant p-values. Many of these associated with autoimmune diseases (psoriasis, lupus, multiple sclerosis).



These are linked SNPs with no assertion of causation. Visit the NOTCH4 gene record and zoom out on the graphical view of that region of chromosome 6. You can see that the NOTCH4 gene is adjacent to the region containing some of the genes (C4B, C4A, CFB) that are part of the complement cascade involved in innate immunity.



Another section of the report shows the dbGaP studies that have asthma as a measured variable. You would need to apply to get access to the individual level data.

**Independent exercise: finding association results for serum triglycerides**

- Use PheGeni to identify the most significant SNP for triglycerides.
- Identify the gene containing or closest to this SNP. Is this gene involved in handling lipids?
- Identify the nearest gene clearly with a role in transporting lipids. Does this gene have a pathogenic variant related to high lipid levels?

Lactase persistence in MedGen, ClinVar, 1000Genomes

Search MedGen for “Lactose Intolerance” as a Clinical Feature.

MedGen MedGen Lactose intolerance Search

Save search Limits Advanced Help

See MedGen results with Lactose intolerance as a clinical feature (2)

Display Settings: Summary 20 per page Send to:

**Results: 16**

- Lactose intolerance**
- 1. An inability to digest lactose. [from HPO]
 

MedGen UID: 50568 • Concept ID: C0004237 • Finding

GTR ClinVar Genes OMIM GeneReviews
- Lactose Intolerance**
- 2. **Lactose intolerance** means that you cannot digest foods with lactose in them. Lactose is the sugar found in milk and foods made with milk. After eating foods with lactose in them, you may feel sick to your stomach. You may also have: - Gas - Diarrhea - Swelling in your stomach . Your doctor may do a blood, breath or stool test to find out if your problems are due to **lactose intolerance**. **Lactose intolerance** is not serious. Eating less food with lactose, or using pills or drops to help you digest lactose usually helps. You may need to take a calcium supplement if you don't get enough of it from your diet, since milk and foods made with milk are the most common source of calcium for most people. NIH: National Institute of Diabetes and Digestive and Kidney Diseases. [from MedlinePlus]
 

MedGen UID: 6001 • Concept ID: C0022951 • Disease or Syndrome

GTR ClinVar Genes OMIM GeneReviews

Filter your results:

- All (16)
- Records in GTR (5)
- Records in OMIM (4)
- Diseases (13)
- Records in Orphanet (3)
- Records in HPO (1)

Manage Filters

Find related data

Database: Select Find items

Search details

lactose intolerance[All Fields]

MedGen MedGen "Lactose intolerance"[Clinical Features] Search

Save search Limits Advanced Help

Display Settings: Summary Send to:

**Results: 2**

- Nonpersistence of intestinal lactase**
- 1. In humans, the activities of lactase and most of the other digestive hydrolases are maximal at birth. The majority of the world's human population experiences a decline in production of the digestive enzyme lactase-phlorizin hydrolase during maturation, with the age of onset ranging from the toddler years to young adulthood. Due to the reduced lactase level, lactose present in dairy products cannot be digested in the small intestine and instead is fermented by bacteria in the distal ileum and colon. The fermentative products result in symptoms of diarrhea, gas bloat, flatulence, and abdominal pain. However, in a minority of adults, high levels of lactase activity persist in adulthood. Lactase persistence is a heritable autosomal dominant condition that results in a sustained ability to digest the milk sugar lactose throughout adulthood (Olds and Sibley, 2003). [from OMIM]
 

MedGen UID: 75659 • Concept ID: C0268181 • Disease or Syndrome

GTR ClinVar Genes OMIM GeneReviews
- Congenital lactase deficiency**
- 2. Congenital lactase deficiency is a severe gastrointestinal disorder characterized by watery diarrhea in infants fed with breast milk or other lactose-containing formulas. [from OMIM]
 

MedGen UID: 120617 • Concept ID: C0268179 • Disease or Syndrome

GTR ClinVar Genes OMIM GeneReviews

Filter your results:

- All (2)
- Records in GTR (2)
- Records in OMIM (2)
- Diseases (2)
- Records in Orphanet (1)
- Records in HPO (0)

Manage Filters

Find related data

Database: Select Find items

Search details

"Lactose intolerance"[Clinical Features]

Search See more...

Retrieve the record for “Nonpersistence of intestinal lactase”.

**Display Settings:**  Full Report **Send to:**

**Nonpersistence of intestinal lactase**  
 MedGen UID: 75659 • Concept ID: C0268181 • Disease or Syndrome

**Synonyms:** ADULT LACTASE DEFICIENCY; DISACCHARIDE INTOLERANCE III; HYPOLACTASIA, ADULT TYPE; Lactose intolerance, adult type

**Modes of inheritance:** Autosomal recessive inheritance

**SNOMED CT:** Ontogenic late onset lactase deficiency (38032004); Non-persistence of intestinal lactase (38032004); Adult lactase deficiency (38032004); Late onset lactase deficiency (38032004); Nonpersistence of intestinal lactase (38032004); Delayed-onset isolated lactase deficiency (38032004); Late-onset lactose intolerance (38032004); Disaccharide intolerance III (38032004); Primary hypolactasia (38032004)

**Gene:** MCM6 

**Cytogenetic location:** 2q21.3

**OMIM®:** 223100

**Definition** **Go to:**

In humans, the activities of lactase and most of the other digestive hydrolases are maximal at birth. The majority of the world's human population experiences a decline in production of the digestive enzyme lactase-phlorizin hydrolase during maturation, with the age of onset ranging from the toddler years to young adulthood. Due to the reduced lactase level, lactose present in dairy products cannot be digested in the small intestine and instead is fermented by bacteria in the distal ileum and colon. The fermentative products result in symptoms of diarrhea, gas bloat, flatulence, and abdominal pain. However, in a minority of adults, high levels of lactase activity persist in adulthood. Lactase persistence is a heritable autosomal dominant condition that results in a sustained ability to digest the milk sugar lactose throughout adulthood (Olds and Sibley, 2003). [from OMIM]

**Additional description** **Go to:**

**Table of contents**

- Definition
- Additional description
- Clinical features
- Recent clinical studies

---

**Genetic Testing Registry**

- Deletion/duplication analysis (2)
- Sequence analysis of select exons (3)
- Sequence analysis of the entire coding region (4)
- Targeted variant analysis (6)

---

See all (13)

---

**Molecular resources**

- OMIM
- RefSeqGene
- View MCM6 variations in ClinVar
- Coriell Institute for Medical Research

MedGen is a largely automatic aggregator of information on phenotypes. It includes several controlled vocabularies for diseases / disorders including SNOMED CT, MeSH, and the vocabulary used by GTR. Notice that this condition is labeled with the MCM6 gene rather than the lactase gene. If you look at the references in the Etiology section of the record, you'll see that the variant involved apparently affects the promoter of the lactase gene, which contains portions of the upstream MCM6 gene.

**Recent clinical studies** **Go to:**

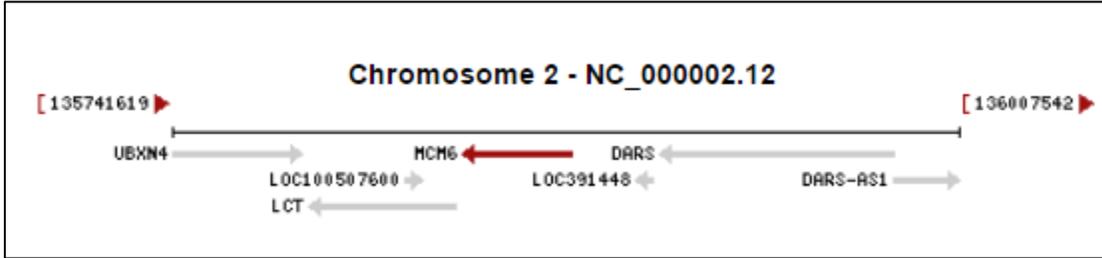
**Etiology**

[Functional significance of single nucleotide polymorphisms in the lactase gene in diverse US patients and evidence for a novel lactase persistence allele at -13909 in those of European ancestry.](#)  
 Baffour-Awuah NY, Fleet S, Montgomery RK, Baker SS, Butler JL, Campbell C, Tischfield S, Mitchell PD, Allende-Richter S, Moon JE, Fishman L, Bousvaros A, Fox V, Kuokkanen M, Grand RJ, Hirschhorn JN  
*J Pediatr Gastroenterol Nutr* 2015 Feb;60(2):182-91. doi: 10.1097/MPG.0000000000000595. PMID: 25625576 Free PMC Article

[Effects of exogenous lactase administration on hydrogen breath excretion and intestinal symptoms in patients presenting lactose malabsorption and intolerance.](#)  
 Ibba I, Gilli A, Boi MF, Usai P  
*Biomed Res Int* 2014;2014:680196. Epub 2014 May 25 doi: 10.1155/2014/680196. PMID: 24967391 Free PMC Article

[The human lactase persistence-associated SNP -13910\\*T enables in vivo functional persistence of lactase promoter-reporter transgene expression.](#)  
 Fang L, Ahn JK, Wodziak D, Sibley E  
*Hum Genet* 2012 Jul;131(7):1153-9. Epub 2012 Jan 19 doi: 10.1007/s00439-012-1140-z. [Epub ahead of print] PMID: 22258180 Free PMC Article

Click through to the MCM6 gene record and navigate to the Genomic Context section of the record. You can see that the lactase gene (LCT) is just to the 3' side of MCM6 and in the same orientation.



Go back to the MedGen record and click the link to ClinVar under Molecular Resources.

The screenshot shows a 'Go to:' dropdown menu with 'Molecular resources' selected. Below the menu, a list of links is displayed: OMIM, View MCM6 variations in ClinVar, RefSeqGen, and Coriell Institute for Medical Research. A blue arrow points upwards from the bottom of the list towards the 'View MCM6 variations in ClinVar' link. On the left side of the screenshot, there is a partial view of text: 'milk and other dairy', 'is produced by cells in', 'enital alactasia, is a', 'mula. This form of lactose', 'e-free infant formula, they', 'thood is caused by', and 'als with lactose'.

	Variation Location	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)
1.	<input type="checkbox"/> <a href="#">MCM6_EX17_T/C</a>	<a href="#">MCM6</a>	Lactate persistence		Pathogenic (Jan 1, 2008)
2.	<input type="checkbox"/> <a href="#">NM_005915.5(MCM6):c.1917+329</a> <a href="#">C&gt;G</a> GRCh37: Chr2:136608643 GRCh38: Chr2:135851073	<a href="#">MCM6</a>	Lactase persistence		association (Jun 16, 2015)
3.	<input type="checkbox"/> <a href="#">NM_002299.2(LCT):c.-13907C&gt;T</a> GRCh37: Chr2:136608646 GRCh38: Chr2:135851076	<a href="#">MCM6</a>	Lactase persistence	GMAF:0.16130(A)	association (Jun 16, 2015)
4.	<input type="checkbox"/> <a href="#">NM_005915.5(MCM6):c.1917+321</a> <a href="#">T&gt;G</a> GRCh37: Chr2:136608651 GRCh38: Chr2:135851081	<a href="#">MCM6</a>	Lactase persistence	GMAF:0.00060(C)	association (Jun 16, 2015)
5.	<input type="checkbox"/> <a href="#">NM_005915.5(MCM6):c.1917+226</a> <a href="#">G&gt;C</a> GRCh37: Chr2:136608746 GRCh38: Chr2:135851176	<a href="#">MCM6</a>	Lactase persistence	GMAF:0.00340(G)	association (Jun 16, 2015)
6.	<input type="checkbox"/> <a href="#">NM_005915.5(MCM6):c.1362+117</a> <a href="#">G&gt;A</a> GRCh37: Chr2:136616754 GRCh38: Chr2:135859184	<a href="#">MCM6</a>	Lactase persistence	GMAF:0.16330(T)	association (Jun 16, 2015)

Notice that four two of these have significant global minor allele frequencies and that one of these is given a location relative to the lactase gene (NM\_002299.2(LCT):c.-13907C>T). This is the common variant that determine lactase persistence in European populations. Retrieve this record.

Home	About ▾	Data use and maintenance ▾	Using the website ▾	How to submit ▾	Statistics
------	---------	----------------------------	---------------------	-----------------	------------

**NM\_002299.2(LCT):c.-13907C>T**

**NM\_002299.2(LCT):c.-13907C>T** Go to:

Variant type: single nucleotide variant

Cytogenetic location: 2q21.3

Genomic location: Chr2:135851076 (on Assembly GRCh38)  
Chr2:136608646 (on Assembly GRCh37)

Other names: MCM6:c.1917+326C>T  
IVS13, C/T

HGVS: NG\_008958.1:g.30366C>T  
NG\_008104.2:g.9094C>T  
NM\_002299.2:c.-13907C>T  
NM\_005915.5:c.1917+326C>T  
NC\_000002.12:g.135851076G>A (GRCh38)  
NC\_000002.11:g.136608646G>A (GRCh37)  
[...less](#)

Links: OMIM: [601806.0001](#)  
dbSNP: [4988235](#)

NCBI 1000 Genomes Browser: [rs4988235](#)

Molecular consequence: NM\_005915.5:c.1917+326C>T: intron variant [Sequence Ontology [SO:0001627](#)]

Allele frequency: GMAF 0.16130 (A)

**NM\_002299.2(LCT):c.-13907C>T**

Variation ID: [?](#) 7685  
 Review status: [?](#) ★ ★ ★ ★ (0/4) no assertion criteria provided

**Interpretation** [?](#) Go to: [v](#) [^](#)

Clinical significance: [association](#)  
 Last evaluated: Jun 16, 2015  
 Number of submission(s): 1  
 Condition(s): Lactase persistence [\[MedGen\]](#)  
[See supporting ClinVar records](#) [↗](#)

**Allele(s)** [?](#) Go to: [v](#) [^](#)

**NM\_002299.2(LCT):c.-13907C>T**

Allele ID: 22724  
 Variant type: single nucleotide variant  
 Cytogenetic location: 2q21.3  
 Genomic location:
 

- Chr2: 135851076 (on Assembly GRCh38)
- Chr2: 136608646 (on Assembly GRCh37)

 Other names:
 

- -13910C\*T
- MCM6:c.1917+326C>T
- IVS13, C/T

 HGVS:
 

- NG\_008958.1:g.30366C>T
- NM\_002299.2:c.-13907C>T
- NM\_005915.5:c.1917+326C>T

[...more](#)

Links:
 

- OMIM: [601806.0001](#)
- dbSNP: [4988235](#)

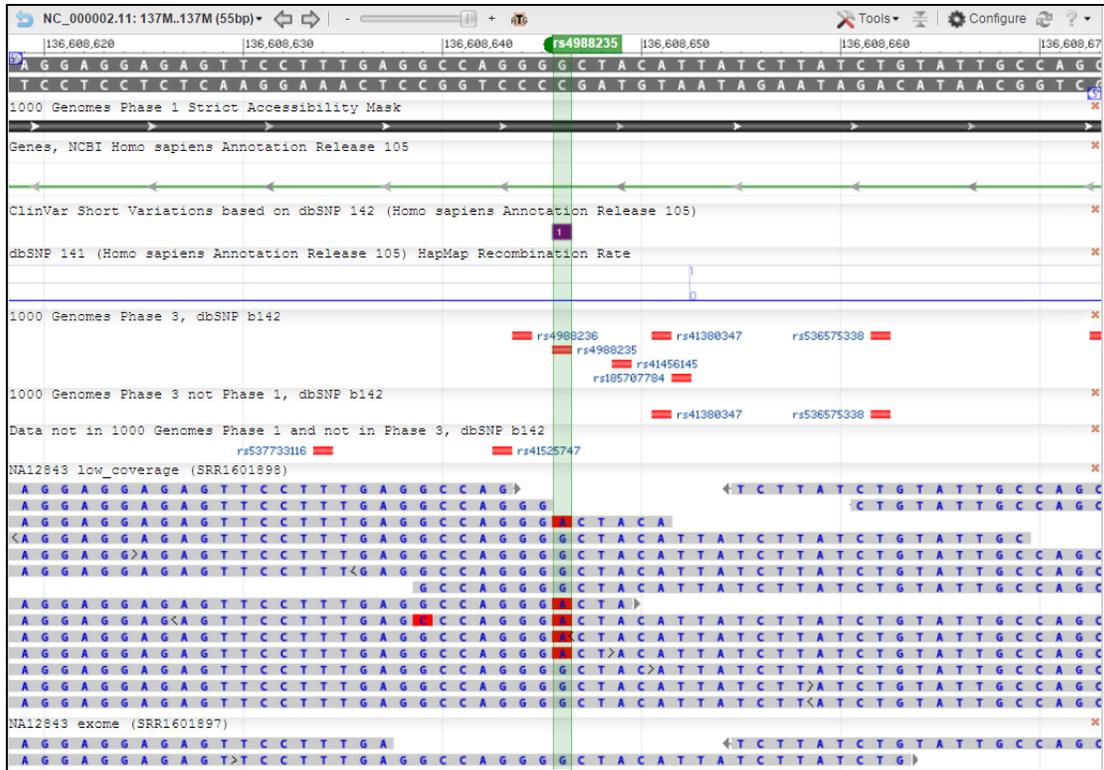
 NCBI 1000 Genomes Browser: [rs4988235](#)  
 Molecular consequence: NM\_005915.5:c.1917+326C>T: intron variant [\[Sequence Ontology SO:0001627\]](#)  
 Allele frequency: GMAF 0.16130 (A)

The main title of this entry is an HGVS expression relative to the CDS start of the lactase (LCT) transcript. The notation indicates position of the variant is 13,907 bases upstream from the start codon of lactase coding region. The HGVS expression also shows that the reference sequence transcript has the “C”, which is the non-persistence allele. Since the gene is on the opposite strand of chromosome 2 the genomic HGVS expressions for chromosome 2 shows the complementary bases (G>A). This indicates that the reference genome assembly shows the non-persistence allele, the more common one worldwide. However the persistence allele is a common variant in some populations. The Global Minor Allele Frequency (GMAF) is nearly 16%. This statistic is from the 1000Genomes dataset. Follow the link to the 1000 Genomes Browser from the ClinVar record to see allele frequencies for human populations. In this case the alleles are reported in the genome context. Thus the persistence allele is an “A”, non-persistence a “G”. The persistence allele

actually is the major allele in populations from northern Europe – greater than 70% in the Utah European ancestry (CEU) and British populations (GBR), 59% in Finnish (Fin), but absent from the Asian populations. The persistence allele is thought to have increased in frequency because of selection pressure from dependence on milk as a food source in northern Europe.

Go to Selection	Scroll Region	136,608,503 rs558877131	136,608,515 rs527991977	136,608,519 rs187602841	136,608,536 rs144412793	136,608,537 rs531916956	136,608,644 rs4988236	136,608,646 rs4988235	136,608,649 rs41456145
<b>Populations / Samples</b>		T=0.9998 C=0.0002	C=0.9990 G=0.0010	C=0.9998 T=0.0002	C=0.9990 A=0.0010	G=0.9996 A=0.0004	G=0.9992 A=0.0008	G=0.8387 A=0.1613	A=0.9998 G=0.0002
▶ ACB	African Caribbeans ...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.9323 A=0.0677	A=1.0000 G=0.0000
▶ ASW	Americans of African...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.8279 A=0.1721	A=1.0000 G=0.0000
▶ BEB	Bengali from Banglad...	T=1.0000 C=0.0000	C=0.9942 G=0.0058	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.9419 A=0.0581	A=1.0000 G=0.0000
▶ CDX	Chinese Dai in Xishu...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	A=1.0000 G=0.0000
▶ CEU	Utah Residents (CEPH...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.2626 A=0.7374	A=1.0000 G=0.0000
▶ CHB	Han Chinese in Beijin...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.9854 A=0.0146	G=1.0000 A=0.0000	A=1.0000 G=0.0000
▶ CHS	Southern Han Chinese	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=0.9952 T=0.0048	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	A=1.0000 G=0.0000
▶ CLM	Colombians from Mede...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.6915 A=0.3085	A=1.0000 G=0.0000
▶ ESN	Esan in Nigera	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	A=1.0000 G=0.0000
▶ FIN	Finnish in Finland	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.4091 A=0.5909	A=1.0000 G=0.0000
▶ GBR	British in England a...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.2802 A=0.7198	A=1.0000 G=0.0000
▶ GIH	Gujarati Indian from...	T=1.0000 C=0.0000	C=1.0000 G=0.0000	C=1.0000 T=0.0000	C=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.8592 A=0.1408	A=1.0000 G=0.0000

Expand the Utah population and add the track for individual NA12843, who is a heterozygote. You can now see the aligned reads in the browser.



These are aligned reads from SRA. Human reads like these are stored in cSRA format that stores only differences from a reference genome. You can also load aligned data directly from the run browser in SRA. To see an example, retrieve experiment SRX461252 from SRA.

**SRX461252: HGDP01259; genome sequencing**  
 1 ILLUMINA (Illumina HiSeq 2000) run: 109.3M spots, 21.7G bases, 9.8Gb downloads

**Accession:** SRX461252  
**Experiment design:** Sample genomic DNA was extracted from lymphoblastoid cell lines and sequenced on an Illumina HiSeq 2000  
**Submitted by:** Stanford University  
**Study summary:** [SRP036155](#) • Transcriptome Sequencing from Diverse Human Populations Reveals Differentiated Regulatory Architecture • [PRJNA236787](#) • [All experiments](#) • [Run Selector \(more...\)](#)  
**Sample:** [SAMN02603825](#) • SRS550849 [\(less...\)](#)  
*Organism:* [Homo sapiens](#)  
*Attributes:*  
 label: HGDP01259  
 Gender: M  
 Population: Mozabite  
 Geographic\_origin: Algeria (Mzab)  
 Geographic\_area: Northern Africa  
 BioSampleModel: Generic

**Library:** [\(more...\)](#)  
**Platform:** Illumina [\(more...\)](#)  
**Pipeline:**

Program	Version
BWA	0.5.9

**Spot descriptor:**

**Total:** 1 run, 109.3M spots, 21.7G bases, [9.8Gb](#) ⓘ

#	Run	# of Spots	# of Bases	Size	Published
1.	<a href="#">SRR1157057</a>	109,263,054	21.7G	<a href="#">9.8Gb</a>	2014-08-26

This is Next-Gen sequencing of genomic DNA from a Mozabite individual from Algeria. This is part of another study of human populations (PRJNA236787). Click the Run link to enter the SRA Run Browser and click on the Alignment tab. Choose chromosome 2 and choose “same sample” as the scope.

**HGDP01259; genome sequencing (SRR1157057)**

Metadata **Alignment** Reads Download

Alignment	Reads	Bases	Fraction
Primary	216.7M	21.6Gbp	99.17%

Reference Range

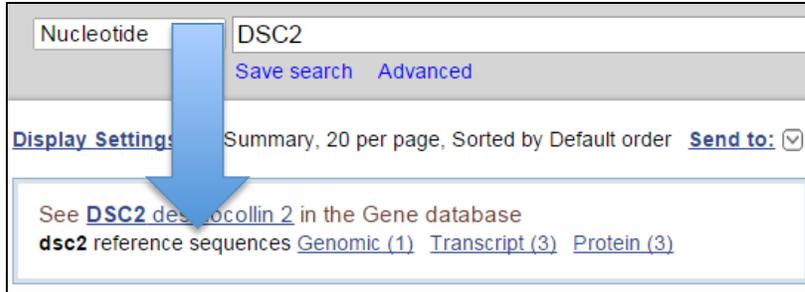
chr2 1-1000000

[Homo sapiens chromosome 2, GRCh37 primary reference assembly](#)  
[What does it do?](#)

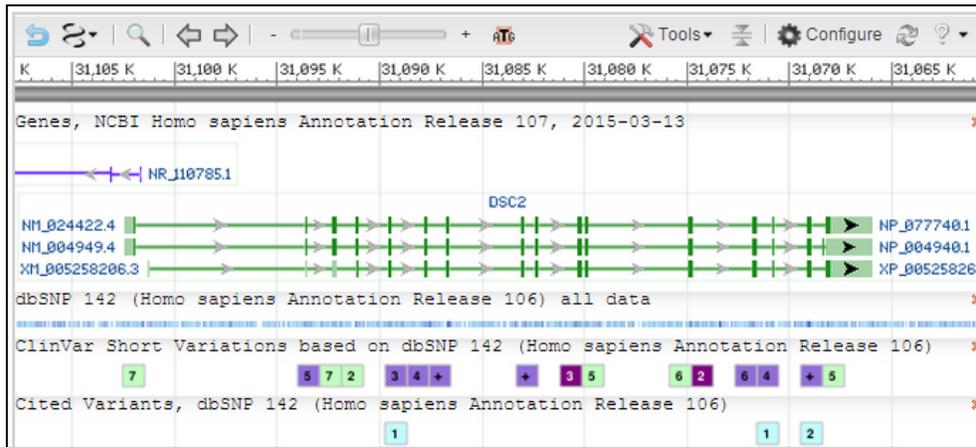
View	scope	accession	count	in
<input type="radio"/>	this run	SRR1157057	1	Sequence Viewer
<input type="radio"/>	same experiment	SRX461252	1	
<input checked="" type="radio"/>	same sample	SRS550849	2	
<input type="radio"/>	same study	SRP036155	107	
<input type="radio"/>	all sra		74,795	

Now click the “Sequence Viewer” button to load the aligned reads. Search for the position of the lactase persistence SNP on chromosome 2 in GRCh37 (136,608,646), then zoom to the sequence at the marker.





The embedded graphical sequence viewer shows a number of pathogenic variants for this gene.



The Phenotypes section of the record suggests the dangerous nature of some of these as there is an ACMG guideline to report about the presence of these variants even as incidental findings.

**Phenotypes**

[Find tests for this gene in the NIH Genetic Testing Registry \(GTR\)](#)

Professional guidelines

Description

Professional guideline

ACMG 2013

The ACMG recommends that laboratories performing clinical sequencing seek and report mutations in DSC2 that are pathogenic or expected to be pathogenic.

[Guideline](#) [PubMed](#)

Associated conditions

Description	Tests
<a href="#">Arrhythmogenic right ventricular cardiomyopathy, type 11</a> MedGen: <a href="#">C1864850</a> , OMIM: <a href="#">610476</a> GeneReviews: <a href="#">Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy</a>	<a href="#">Compare labs</a>

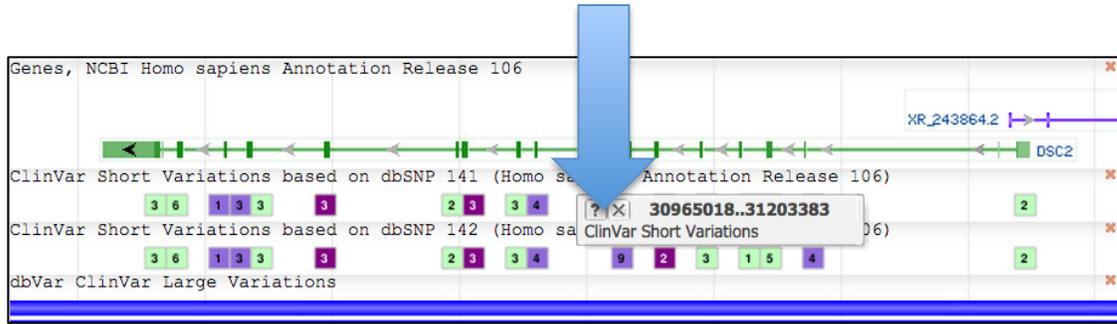
We will use the variation viewer, a dedicated browser for exploring all variants mapped to a region of the genome, to explore variants in this gene and other nearby regions. It includes both large variants from dbVar and small variants from dbSNP.

### Accessing the Variation Viewer

Follow the link from the DSC2 gene record to the Variation Viewer.

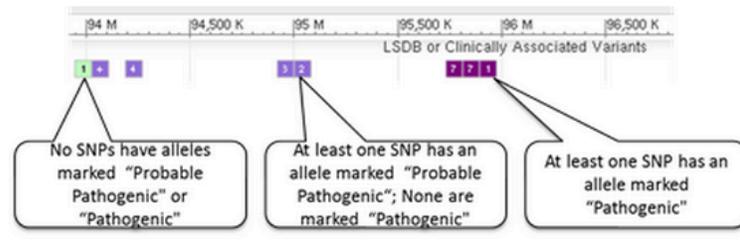
### Tracks in the sequence viewer display

Notice the main graphical display. At the very top is an overview of human chromosome 18 represented as an ideogram with cytobands shown. There is a marker on the ideogram showing where we are on chromosome 18. We could drag this to move the display to another location. Below the ideogram is a navigation device. This is the Gene/Exon navigator. We'll use this to move around within the gene region in a few minutes. The graphic display below this is the embedded NCBI graphical sequence viewer that we saw in gene and in the 1000 Genomes Browser. The sequence viewer shows the region of chromosome 18 that contains the DSC2 gene. There are several tracks displayed. At the very top is an unlabeled track represented by a gray bar that is the sequence of chromosome 18. The next track shows the annotated genes. In this case DSC2. In this view the gene graphic has all splice variants merged but you can still see the exon intron structure. The next track shows the locations of small-scale variants from SNP that are in ClinVar and therefore have assertions about their phenotypes. The number in the boxes represents the number of variants in the region cover by the box. The color tells about pathogenicity. Put your mouse pointer over the gray area at the top of the track and link to the "Track legend" so you can see how the color code works.



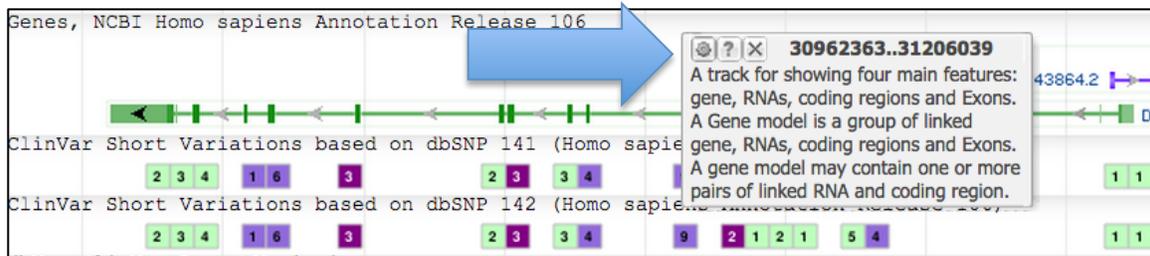
**4.3. SNP Bins For Clinical Associations**

Color	Description
Light Green	No SNPs in this bin have an allele marked "Probable Pathogenic" or "Pathogenic"
Light Purple	At least one SNP in this bin has an allele marked "Probable Pathogenic"; none are "Pathogenic"
Purple	At least one SNP in this bin has an allele marked "Pathogenic"

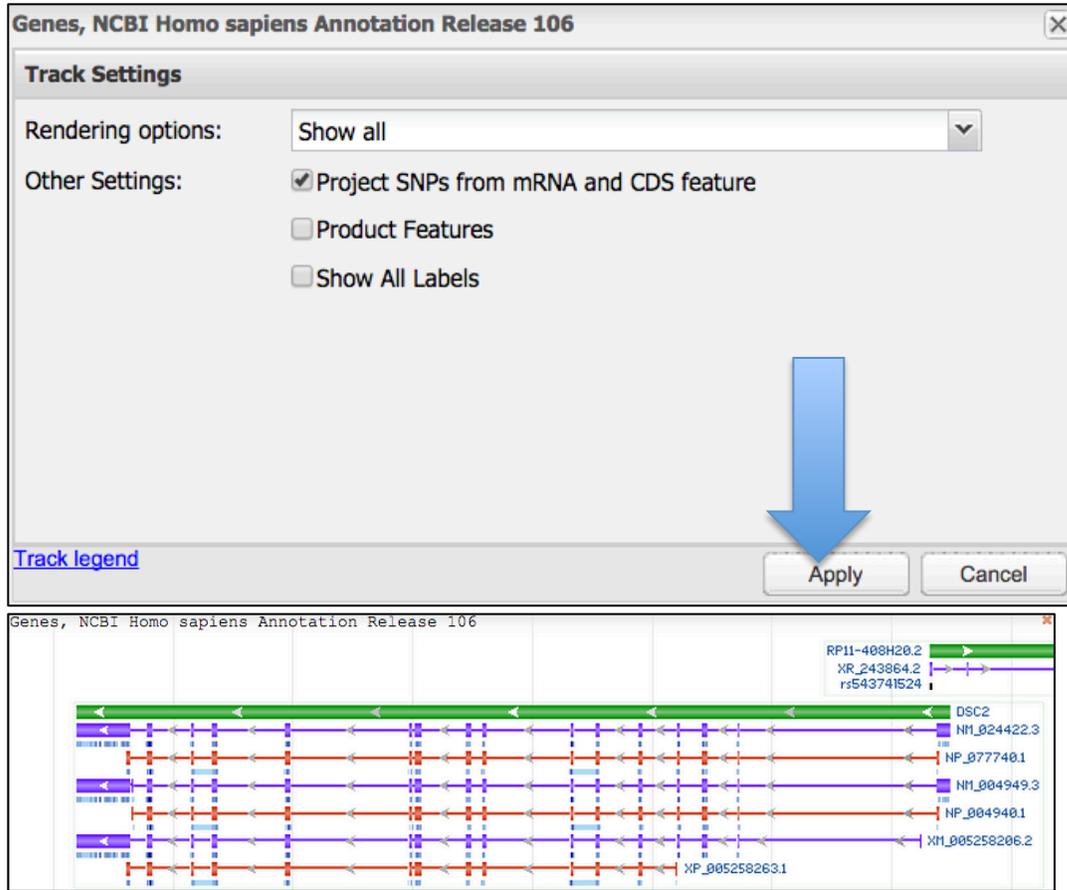


The next track shows large-scale variant regions from dbVar that have clinical assertions in ClinVar. These cover many genes. The red ones are relative deletions. The blue ones are duplications. You can check this with the "Track legend". The track at the bottom shows the locations of all variants from SNP. At this scale the track seems like nearly continuous data.

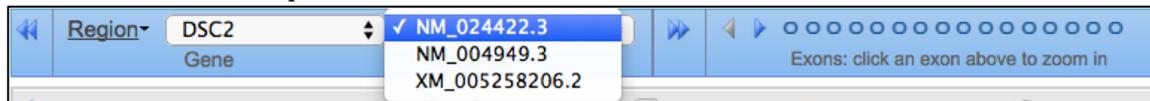
Change the Genes display to in the Variation Viewer to show all of the splice variants. Mouse over the top of the track and click the gear icon.



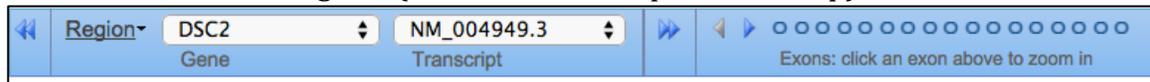
The Configure dialog appears for the Genes track. Set "Show all" in the rendering options and also check the "Project SNPs from mRNA and CDS feature", then click configure to apply these.



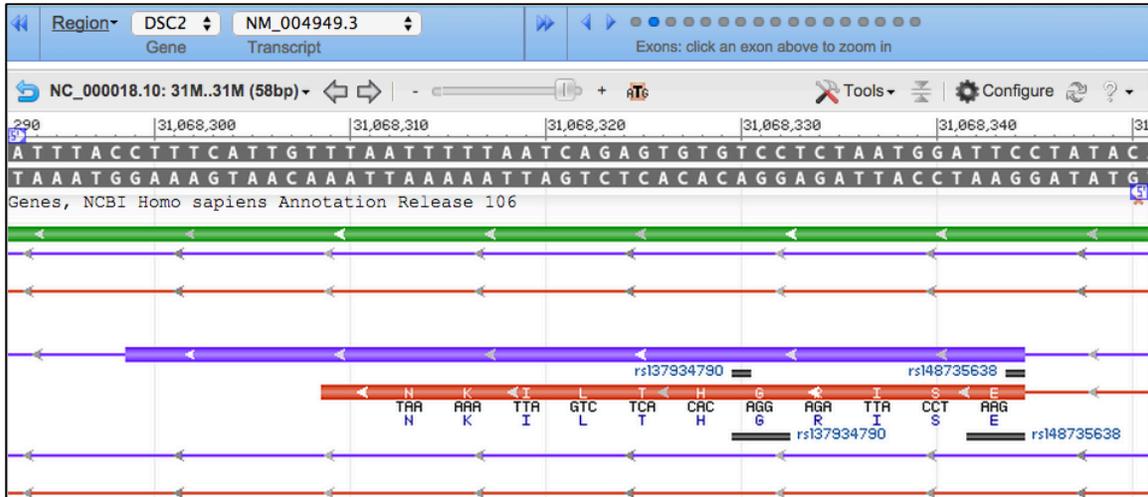
We can now see that there are three splice variants for this gene with slightly different sets of exons. We can use the Gene/Exon navigator to jump to the exon for the different transcript.



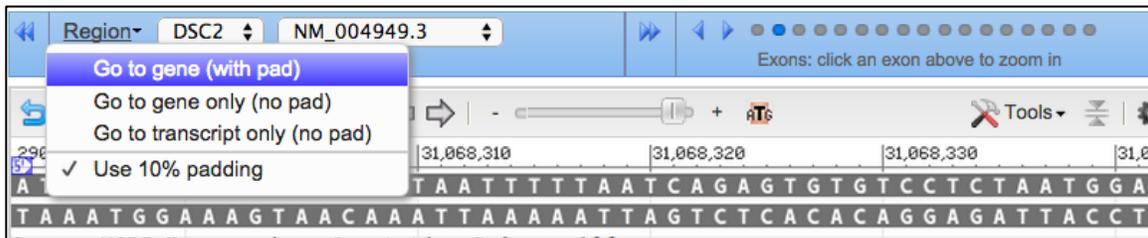
For example if I select NM\_004949 as the transcript, notice that the number of exons shown in the exon navigator (the radio button spots at the top) increases.



Jump to the second to the last exon by clicking on the second spot from the left. This is the second to the last exon in this case. Notice that this gene is on the opposite strand of the chromosome, which is why the 3-prime end appears at the left. This small exon only occurs in this splice variant and not in the other two.



You are now zoomed in to the sequence level and can see nucleotides and amino acid sequences of the gene, transcript and products as well as individual SNP positions. You can restore the view using the Exon/Gene Navigator to select “Go to gene (with pad)” from the “Region” menu.



### Region details and filters

Shift your focus to the left hand side of the Variation Viewer again. Starting at the top notice that we are currently showing the Genome Reference Consortium build 38 with NCBI’s current annotation. The Genome Reference Consortium is a multi-center effort that includes NCBI responsible for maintaining updating and improving a reference assembly of the human genome. This is the descendant of the original publically Human Genome project sequence. You could also choose to show build 37 the previous genome assembly.



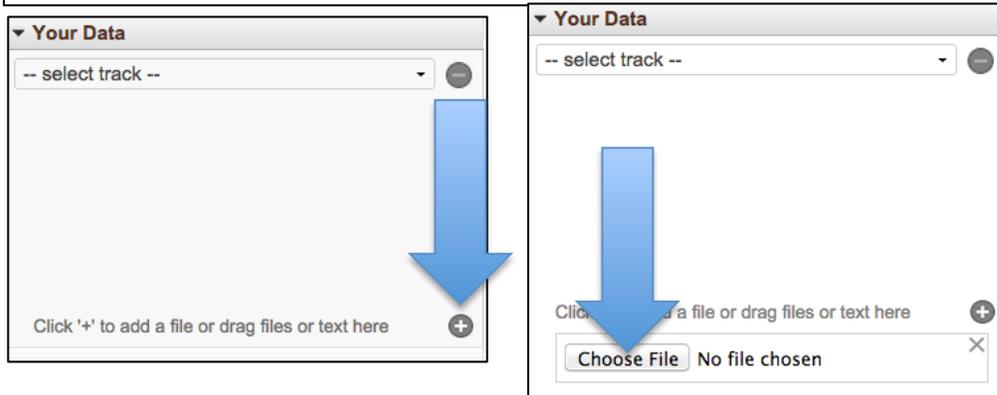
The “Region details” is where you may see other sequence representations such as alternate assemblies of the region representing other haplotypes if these were available. You may also see known issues with the current GRC build. In this case

there are issues in the region. You can use the short cut link to add the track to the display. This is the first track you've added. You can add others through the sequence viewer configure dialog. Going back to the sequence viewer display for a moment, drag the newly added track down below the Sequence track and mouse over the black graphic for the GRC issue to see that there is a problem with the sequence of the clone chosen for this region of the assembly. This is an issue that the GRC will resolve and of necessary make a patch to be incorporated in the next genome build.

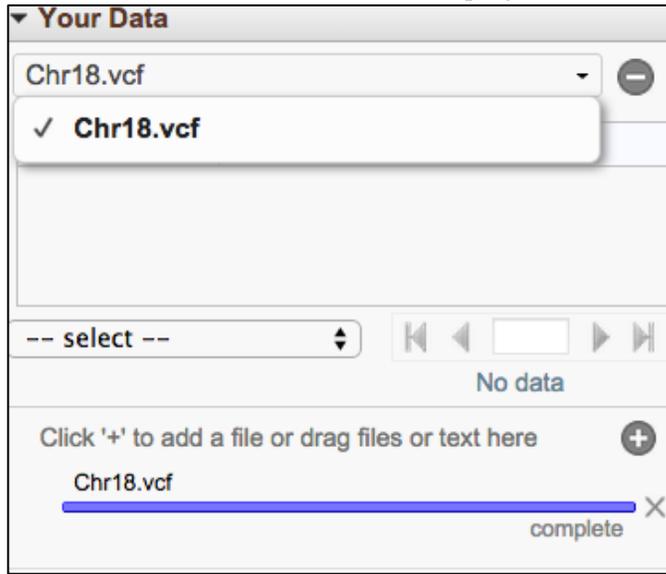
### Uploading your own data

Before we leave this area of the viewer, notice that you can load your own data into the Variation Viewer. You'll need a file with your variant calls. The upload feature accepts a number of different formats including BED, HGVS, GVF and VCF. You can then use the "Your Data" section to upload your data. A minimalist VCF file (Chr18.vcf) is given below. You can load it in your Web browser though the following link: <http://1.usa.gov/1xMOt6r>. You can save it as a text file and upload it as a track.

```
##fileformat=VCFv4.0
##reference=GRCh38
#CHROM      POS      ID       REF      ALT      QUAL    FILTER    INFO
18      31074696  Snp1    C        A        .        .        My=Snp1
18      31074750  Snp2    G        A        .        .        My=Snp2
18      31074875  Snp11   T        C        .        .        My=Snp11
18      31074892  Snp12   G        A        .        .        My=Snp12
18      31079872  Snp3    A        G        .        .        My=Snp3
18      31079958  Snp4    C        G        .        .        My=Snp4
18      31096565  Snp6    C        CG       .        .        My=Snp6
18      31086483  Snp3    C        A        .        .        My=Snp3
```



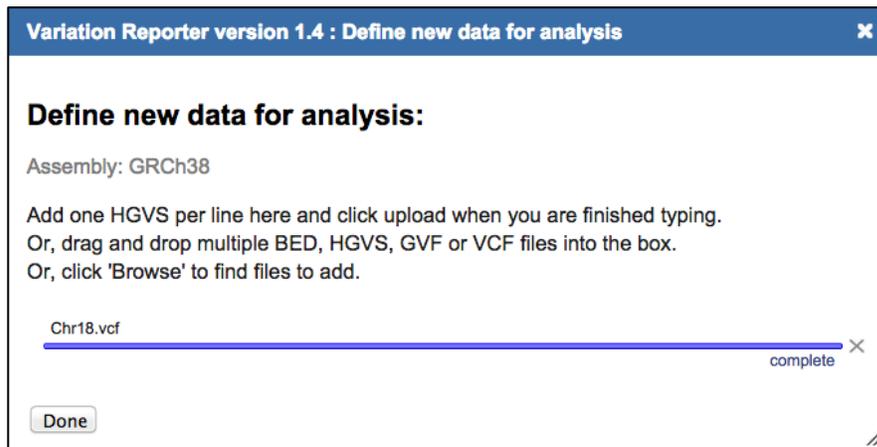
Select the track to add it to the display.



dbSNP 142 (Homo sapiens Annotation Release 106) all data					
Chr18.vcf			Snp1   Snp2   Snp11   Snp12	Snp3   Snp4	Snp3   Snp6

*The Variation Reporter*

A related resource is the Variation Reporter. It provides functional consequences for a set of variants and maps them onto chromosomes, transcripts and proteins. You can also load this file through the Variation Reporter ([www.ncbi.nlm.nih.gov/variation/tools/reporter](http://www.ncbi.nlm.nih.gov/variation/tools/reporter)), which is linked to the Variation Portal page. Download the file to your computer and upload into the Variation Reporter.



**Choose your data context**

Organism:

Assembly:

---

**Your data**

Available data

File name	Track name	Assembly	Select	Delete
Chr18.vcf	Chr18.vcf	GRCh38	<input type="radio"/>	<span style="color: red;">✖</span>

Click on the value in the Submitted Loc column to show it in Sequence Viewer. [Download Report](#)

Items 1 - 10 of 24 << First < Prev Page 1 of 3 Next > Last >>

Submitted Id	Submitted Loc	Allele	Cytoband	NCBI Id	Origin	1000Genomes MAF	Clinical Information	PubMed	Transcript Allele	Consequences
Snip1	<a href="#">NC_000018.10:31074696</a>	NC_000018.10:g.31074696C>A	18q12.1	<a href="#">rs142594406</a>					XM_005258206.2:c.1446G>T	<a href="#">synonymous_c</a>
Snip1	<a href="#">NC_000018.10:31074696</a>	NC_000018.10:g.31074696C>A	18q12.1	<a href="#">rs142594406</a>					NM_004949.3:c.1875G>T	<a href="#">synonymous_c</a>
Snip1	<a href="#">NC_000018.10:31074696</a>	NC_000018.10:g.31074696C>A	18q12.1	<a href="#">rs142594406</a>					NM_024422.3:c.1875G>T	<a href="#">synonymous_c</a>
Snip2	<a href="#">NC_000018.10:31074750</a>	NC_000018.10:g.31074750G>A	18q12.1						XM_005258206.2:c.1392C>T	<a href="#">synonymous_c</a>
Snip2	<a href="#">NC_000018.10:31074750</a>	NC_000018.10:g.31074750G>A	18q12.1						NM_004949.3:c.1821C>T	<a href="#">synonymous_c</a>
Snip2	<a href="#">NC_000018.10:31074750</a>	NC_000018.10:g.31074750G>A	18q12.1						NM_024422.3:c.1821C>T	<a href="#">synonymous_c</a>
Snip11	<a href="#">NC_000018.10:31074875</a>	NC_000018.10:g.31074875T>C	18q12.1						XM_005258206.2:c.1267A>G	<a href="#">non_synonymc</a>
Snip11	<a href="#">NC_000018.10:31074875</a>	NC_000018.10:g.31074875T>C	18q12.1						NM_004949.3:c.1696A>G	<a href="#">non_synonymc</a>
Snip11	<a href="#">NC_000018.10:31074875</a>	NC_000018.10:g.31074875T>C	18q12.1						NM_024422.3:c.1696A>G	<a href="#">non_synonymc</a>
Snip12	<a href="#">NC_000018.10:31074892</a>	NC_000018.10:g.31074892G>A	18q12.1						XM_005258206.2:c.1250C>T	<a href="#">non_synonymc</a>

Items 1 - 10 of 24 << First < Prev Page 1 of 3 Next > Last >>

The Variation Reporter is particularly useful if you have your own variant calls and want to map them on to the genome and products. Notice that it provides the functional consequences for the variant calls and indicate when they match known SNPs. Follow the link to Snip11 to display it in the graphical sequence viewer at the bottom of the output

**Using Filters and Variant Table**

You can use the Variant Filters at the bottom of the right-hand column to find variants of interest. Scroll through these so you can see the kinds of things you can filter by. For example you can filter by minor allele frequency, and on small-scale variants. Select dbSNP as the first filter and then find common variants. Use the GO-ESP (Grand Opportunity Exome Sequencing Project), a set of variants from a number of large-scale genotyping studies. You could also select allele frequencies

from the 1000Genomes project data, a different set of populations from all over the world. Select minor allele frequency  $\geq 0.05$ .

**Filter by** ?

**Source database**

dbSNP (893)

dbVar (0)

**In ClinVar**

Yes (57)

No (836)

**GO-ESP MAF**

< 0.005 (0)

0.005 - 0.01 (0)

0.01 - 0.05 (0)

$\geq 0.05$  (2)

not specified (0)

This leaves only two variants in the table.

Variant ID	Location	Variant type	Gene	Molecular consequences	Worst clinical significance	1000G MAF	GO-ESP MAF	Publications
rs1893963	<a href="#">31,069,076</a>	single nucleotide variant	DSC2	missense variant	Benign	C = 0.1892	C = 0.1912	2
rs12954874	<a href="#">31,093,602</a>	single nucleotide variant	DSC2	synonymous variant, 5 prime UTR variant	Benign	C = 0.1157	C = 0.1071	

Take a look at the missense variant rs1893963 with a minor allele frequency of 19%. You can link to this variant in the viewer by clicking on the location. Notice that there are overlapping variants in this region, one that affects a different position of in the codon and results in a different amino acid change and one that is redundant and will be merged with rs1893963 in the next dbSNP build.

NC\_000018.10: 31M..31M (53bp)

1,069,050 31,069,060 31,069,070 31,069,080 31,069,090 31,069,100

A T G G T C T C C T G A C C T C C G T T T T G A T T C C T G A T C C C A C G G T G C C A C A A A C T C C

T A C C A G A G G A C T G G A G G C A A A A C T A A G G A C T A G G G T G C C A C G G T G T T T G A G G

Genes, NCBI Homo sapiens Annotation Release 106

rs139290300 rs1789054 rs1893963 rs146029947 rs139558481

I T E O G G N K I S G U T G C U G

CTA CCA GAG GAC TGG AGG CAA AAA CTA AGG ACT AGG GTG CCA CGG TGT TTG AGG

I T E O G G N K I S G U T G C U G

CTA CCA GAG GAC TGG AGG CAA AAA CTA AGG ACT AGG GTG CCA CGG TGT TTG AGG

ClinVar Short Variations based on dbSNP 141 (Homo sapiens Annotation Release 106)

ClinVar Short Variations based on dbSNP 142 (Homo sapiens Annotation Release 106)

Go back to the Exon/Gene navigator and restore the full-view as before by selecting "Go to gene (with pad)" in the "Region" menu. Now you can find some pathogenic

variants using the Variant Filter options. Uncheck the minor allele frequency option and set “In ClinVar” to “Yes”.

**Filter by**  
**Source database**  
 dbSNP (57)  
 dbVar (0)  
**In ClinVar**  
 Yes (57)  
 No (0)

You should now have 57 variants that have information in ClinVar. This may be a set that you’d want to download. You can do this using the download link at the top of the table. This delivers an XML file that has all the information in the table for each of the variants including the expanded information, as I’ll show in a minute. Now check pathogenic under “Worst clinical significance”.

**Source database**  
 dbSNP (3)  
 dbVar (0)  
**In ClinVar**  
 Yes (3)  
 No (0)  
**Worst clinical significance**  
 Pathogenic (3)  
 Likely pathogenic (0)  
 risk factor (0)  
 Uncertain significance (0)  
 not provided (0)

You should get three variants; two that result in frame shifts and one that affects a splice site. Expand one of the frameshift ones, rs397514041.



This variant comes from the OMIM record and hence is labeled as literature only. OMIM is only one of the sources of assertions in ClinVar. Follow the link to MedGen on the right to learn more about this particular syndrome. MedGen aggregates data from number of sources and is particularly helpful for integrating controlled vocabularies for phenotypic terms. (See the Clinical Features and Term Hierarchy section of the record.) Notice that this condition can result in sudden cardiac death. Because of this the ACMG recommends that any incidental finding on certain variants in the DSC2 gene and others be reported to patients undergoing sequencing for other genetic conditions. You can link to that paper from the bottom of the MedGen record. There are links to specific genetic tests in the Genetic Testing Registry here as well.

MedGen C1864850[conceptid] Search

Save search Limits Advanced Help

Display Settings:  Full Report Send to:

**Arrhythmogenic right ventricular cardiomyopathy, type 11 (ARVD11)**  
 MedGen UID: 351237 • Concept ID: C1864850 • Disease or Syndrome

**Synonyms:** ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY 11; Arrhythmogenic right ventricular dysplasia 11; ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL, 11; Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy11; ARVD11

**Modes of inheritance:** Autosomal recessive inheritance  
 Autosomal dominant inheritance

**Gene:** DSC2  
**Cytogenetic location:** 18q12.1  
**OMIM®:** 610476

**Disease characteristics** Go to:

**Excerpted from the GeneReview: Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy**  
 Arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARVD/C) is characterized by progressive fibrofatty replacement of the myocardium that predisposes to ventricular tachycardia and sudden death in young individuals and athletes. It primarily affects the right ventricle; with time, it may also involve the left ventricle. The presentation of disease is highly variable even within families, and some affected individuals may not meet established clinical criteria. The mean age at diagnosis is 31 years (±13; range: 4-64 years). [from GeneReviews]

**Full text of GeneReview (by section):**  
[Summary](#) | [Diagnosis](#) | [Clinical Description](#) | [Differential Diagnosis](#) | [Management](#) | [Genetic Counseling](#) | [Resources](#) | [Molecular Genetics](#) | [References](#) | [Chapter Notes](#)

**Authors:**  
 Elizabeth McNally | Heather MacLeod | Lisa Dellefave-Castillo [view full author information](#)

**Additional description** Go to:

**From GHR**  
 Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a form of heart disease that usually appears in adulthood. ARVC is a disorder of the myocardium, which is the muscular wall of the heart. This condition causes part of the myocardium to break down over time, increasing the risk of an abnormal heartbeat (arrhythmia) and sudden death. ARVC may not cause any symptoms in its early stages. However, affected individuals may still be at risk of sudden death, especially during strenuous exercise. When symptoms occur, they most commonly include a sensation of fluttering or pounding in the chest (palpitations), light-headedness, and fainting (syncope). Over time, ARVC can also cause shortness of breath and abnormal swelling in the legs or abdomen. If the myocardium becomes severely damaged in the later stages of the disease, it can lead to heart failure. <http://ghr.nlm.nih.gov/condition/arrhythmogenic-right-ventricular-cardiomyopathy>

**Table of contents**

- Disease characteristics
- Additional description
- Clinical features
- Term Hierarchy
- Professional guidelines
- Recent clinical studies

**Genetic Testing Registry**

- Deletion/duplication analysis (14)
- Mutation scanning of the entire coding region (1)
- Sequence analysis of select exons (2)
- Sequence analysis of the entire coding region (29)
- See all (30)

**Molecular resources**

- OMIM
- RefSeqGene
- View DSC2 variations in ClinVar
- Coriell Institute for Medical Research

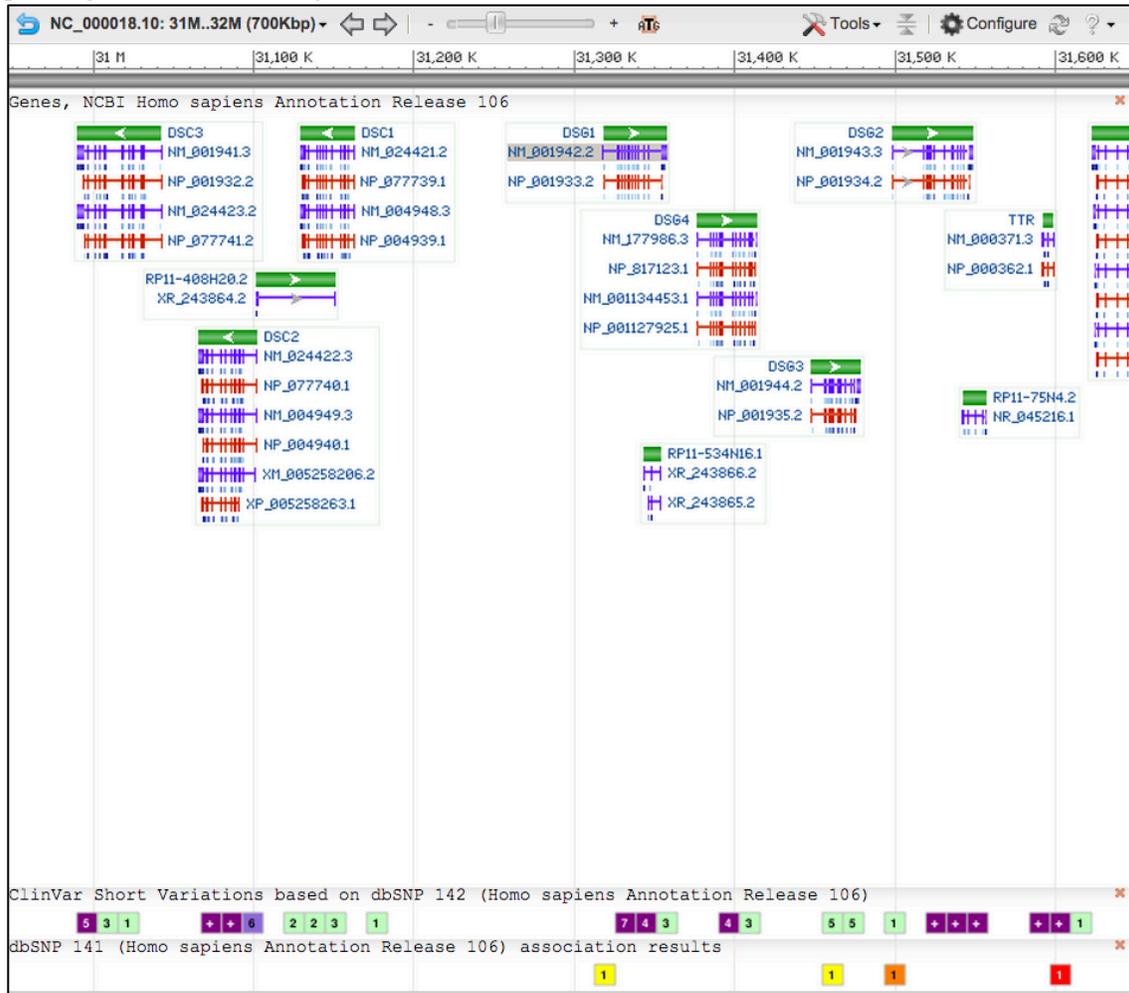
**Consumer resources**

- Genetics Home Reference
- Genetic Alliance
- MedlinePlus

## Adding tracks and navigating

Go back to the Variation Viewer to add an additional track. Click the configure menu in the sequence viewer and select the Variation tracks on the right-hand-side. Check

the box for “dbSNP 141 association results”. This track tells you about genotype-phenotype association results like the ones shown in PheGenI from a number of sources. Click the Configure button to add this track. You will see that there is nothing in this view. Zoom out and move right, down the q arm of chromosome 18. Notice the other desmosomal protein genes in this region, the desmocollins (DSC1, DSC2, DSC3) and the desmogleins (DSG1 through DSG4). Many of these have pathogenic variants as you can see.



### Association results

You can see a highly significant association result just downstream of the TTR gene. These association results are color coded by negative log p value. Red is most significant. Mouse over this downstream result for TTR. Notice that the variant is associated (strongly) with vitamin A levels. This is from a Genome Wide Association Study (GWAS).

The screenshot shows the NCBI genome browser interface for chromosome 10. The top track displays the gene **TTR** (transthyretin) with its location (31,591,767..31,599,024) and length (7,258). Below it, the **dbSNP 141** track shows an association result for variant **rs1667255** at location 31607316..31607316. The variant has a P-value (-log10) of 13.2 and is associated with **Vitamin A**. A blue arrow points from the gene track to the dbSNP track, and another blue arrow points from the dbSNP track to the gene track.

Remember this is an association (linkage) not an assertion that this is a causative variant. Follow the link to gene to that TTR is a retinol (vitamin A) carrier protein, so this is a potentially interesting result.

**TTR transthyretin [ *Homo sapiens* (human) ]**

Gene ID: 7276, updated on 9-Nov-2014

**Summary**

**Official Symbol** TTR provided by [HGNC](#)

**Official Full Name** transthyretin provided by [HGNC](#)

**Primary source** [HGNC:HGNC:12405](#)

**See related** [Ensembl:ENSG00000118271](#); [HPRD:01447](#); [MIM:176300](#); [Vega:OTTHUMG00000131984](#)

**Gene type** protein coding

**RefSeq status** REVIEWED

**Organism** [Homo sapiens](#)

**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

**Also known as** CTS; CTS1; PALB; TBPA; HEL111; HsT2651

**Summary** This gene encodes transthyretin, one of the three prealbumins including alpha-1-antitrypsin, transthyretin and orosomucoid. Transthyretin is a carrier protein; it transports thyroid hormones in the plasma and cerebrospinal fluid, and also transports retinol (vitamin A) in the plasma. The protein consists of a tetramer of identical subunits. More than 80 different mutations in this gene have been reported; most mutations are related to amyloid deposition, affecting predominantly peripheral nerve and/or the heart, and a small portion of the gene mutations is non-amyloidogenic. The diseases caused by mutations include amyloidotic polyneuropathy, euthyroid hyperthyroxinaemia, amyloidotic vitreous opacities, cardiomyopathy, oculoleptomeningeal amyloidosis, meningocerebrovascular amyloidosis, carpal tunnel syndrome, etc. [provided by RefSeq, Jan 2009]

You can link through to PubMed to see that the authors find two significant associations in this article. The other one is near RBP4 (retinol binding protein 4), which also makes some biological sense.

*Hum Mol Genet.* 2011 Dec 1;20(23):4724-31. doi: 10.1093/hmg/ddr387. Epub 2011 Aug 30.

**Genome-wide association study of circulating retinol levels.**

Mondul AM<sup>1</sup>, Yu K, Wheeler W, Zhang H, Weinstein SJ, Major JM, Cornelis MC, Männistö S, Hazra A, Hsing AW, KB, Eliassen H, Tanaka T, Reding DJ, Hendrickson S, Ferrucci L, Virtamo J, Hunter DJ, Chanock SJ, Kraft P, Albers

**Author information**

**Abstract**

Retinol is one of the most biologically active forms of vitamin A and is hypothesized to influence a wide range of human diseases including asthma, cardiovascular disease, infectious diseases and cancer. We conducted a genome-wide association study of 5006 Caucasian individuals drawn from two cohorts of men: the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study and the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. We identified two independent single-nucleotide polymorphisms associated with circulating retinol levels, which are located near the transthyretin (TTR) and retinol-binding protein 4 (RBP4) genes which encode major carrier proteins of retinol: rs1667255 (P =2.30× 10<sup>-17</sup>, rs10882272 (P =1.10× 10<sup>-12</sup>). We replicated the association with rs10882272 in RBP4 in independent samples from the Nurses' Health Study and the Invecchiare in Chianti Study (InCHIANTI) that included 3792 women and 504 men (P =9.49× 10<sup>-5</sup>), but found no association for retinol with rs1667255 in TTR among women, thus suggesting evidence for gender dimorphism (P-interaction=1.31× 10<sup>-5</sup>). Discovery of common genetic variants associated with serum retinol levels may provide further insight into the contribution of retinol and other vitamin A compounds to the development of cancer and other complex diseases.

**Related information**

- Related Citations
- Gene
- HomoloGene
- MedGen
- Nucleotide (RefSeq)
- Nucleotide (Weighted)
- Protein (RefSeq)
- Protein (Weighted)
- PubChem Compound (MeSH Keyword)
- PubChem Substance (MeSH Keyword)
- References for this PMC Article
- SNP (Cited)**
- Taxonomy via GenBank
- UniGene
- GEO Profiles
- Free in PMC
- Cited in PMC

To explore this, link to SNP cited from the “Related information” section on the right-hand side of the PubMed abstract.

**Results: 2**

**rs10882272** [*Homo sapiens*]

1.

- **Suspected**

TTTTTTTTTTCATTTATAAAAATGC [C/T] ATGGACCTTTTAAAGAGAATCGGCA

Chromosome: 10:93588425

Gene: FFAR4 ([GeneView](#))

Functional Consequence: utr variant 3 prime

Validated: no info

Global MAF: C=0.3898/1951

HGVS: NC\_000010.10:g.95348182T>C,  
 NC\_000010.11:g.93588425T>C,  
 NC\_000010.7:g.95012769T>C, NG\_032670.1:g.26761T>C,  
 NM\_001195755.1:c.\*816T>C, NM\_181745.3:c.\*816T>C

[PubMed](#)

**rs1667255** [*Homo sapiens*]

2.

GCCAGAGATGGGACTATTTCTTCTT [A/C] TTGTTTTAGATGTAACATTAAAAA

Chromosome: 18:31607316

Validated: no info

Global MAF: A=0.4996/2502

HGVS: NC\_000018.10:g.31607316A>C,  
 NC\_000018.9:g.29187279A>C

[PubMed](#)

There is no SNP associated with RBP4. Let's see why. Follow the link through to rs10882272 (FFAR4 gene), then from the SNP record link through to build 38 Chr Pos. to the Variation Viewer.

Reference SNP (refSNP) Cluster Report: rs10882272

RefSNP	Allele	HGVS Names
Organism: human ( <i>Homo sapiens</i> )	<b>Variation Class:</b> SNV: single nucleotide variation	NC_000010.10:g.95348182T>C
Molecule Type: Genomic	<b>RefSNP Alleles:</b> C/T (FWD)	NC_000010.11:g.93588425T>C
Created/Updated in build: 120/142	<b>Allele Origin:</b>	NC_000010.7:g.95012769T>C
Map to Genome Build: 106/Weight	<b>Ancestral Allele:</b> C	NG_032670.1:g.26761T>C
<b>Validation Status:</b>	<b>Variation Viewer:</b> unknown	NM_001195755.1:c.*816T>C
<b>Citation:</b> PubMed	<b>Clinical Significance:</b> NA	NM_181745.3:c.*816T>C
<b>Association:</b> NHGRI GWAS PheGenI	<b>MAF/MinorAlleleCount:</b> C=0.3898/1952	
	<b>MAF Source:</b> 1000 Genomes	

SNP Details are organized in the following sections: [GeneView](#) [Map](#) [Submission](#) [FAQ](#) [Resource](#) [Diversity](#) [Validation](#)

Integrated Maps (Hint: click on 'Chr Pos' to view variant in the new NCBI variation viewer)

Assembly	Annotation Release	Chr	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr
GRCh38	106	10	93588425	NT_030059.14	51894904	Fwd T	Fwd
GRCh37.p13	105	10	95348182	NT_030059.13	46152646	Fwd T	Fwd

Zoom out so that you can see rs10882272 is also close to and upstream of RPB4.

Homo sapiens: GRCh38 (GCF\_000001405.26) Chr 10 (NC\_000010.11): 93.56M - 93.62M

Region: FFAR4 (Gene) NM\_181745.3 (Transcript)

NC\_000010.11: 94M..94M (55Kbp)

Genes, NCBI Homo sapiens Annotation Release 106

**FFAR4**

Gene: FFAR4  
Title: free fatty acid receptor 4  
Location: 93,566,665..93,590,072  
Length: 23,408

Links & Tools  
View GeneID: [338557 \(FFAR4\)](#)  
View HGNC: [19061](#)  
View HPRD: [16422](#)  
View MIM: [609044](#)

GenBank View: [NC\\_000010.11 \(93,566,665..93,590,072\)](#)  
FASTA View: [NC\\_000010.11 \(93,566,665..93,590,072\)](#)  
BLAST Genomic: [NC\\_000010.11 \(93,566,665..93,590,072\)](#)

**RBP4**

Gene: RBP4  
Title: retinol binding protein 4, plasma  
Location: complement(93,591,836..93,601,344)  
Length: 9,509

Links & Tools  
View GeneID: [5950 \(RBP4\)](#)  
View HGNC: [9922](#)  
View MIM: [180250](#)

GenBank View: [NC\\_000010.11 \(93,591,836..93,601,344\)](#)  
FASTA View: [NC\\_000010.11 \(93,591,836..93,601,344\)](#)  
BLAST Genomic: [NC\\_000010.11 \(93,591,836..93,601,344\)](#)

### Independent exercise: variants on chromosome 5

Investigate the following set of variants on chromosome 5 using the Variation Reporter, the Variation Viewer, SNP, ClinVar and any other NCBI resources.

NC\_000005.10:g.1254479C>T  
NC\_000005.10:g.1286401C>A  
NC\_000005.10:g.1321972C>T  
NC\_000005.10:g.1421890T>A  
NC\_000005.10:g.1421997A>T

- Where possible match these to RefSNP records.
- Identify the two affected genes.
  - What are the functions of these genes?
- Predict the functional consequences of each on the mRNA and protein sequences.
- Find any phenotypes associated with each.
- Identify any previously unreported variants
- Identify any novel alleles.
- Which two of these have a minor allele frequency greater than 1% in the 1000Genomes populations?
  - Find the genotypes of individual samples NA12878 (CEU population) and NA19240 (YRI Population).
  - Find any GWAS associations for the polymorphic SNPs. What conditions are linked to these?
- Which of the variants have information in ClinVar?
  - What conditions reported for these variants?