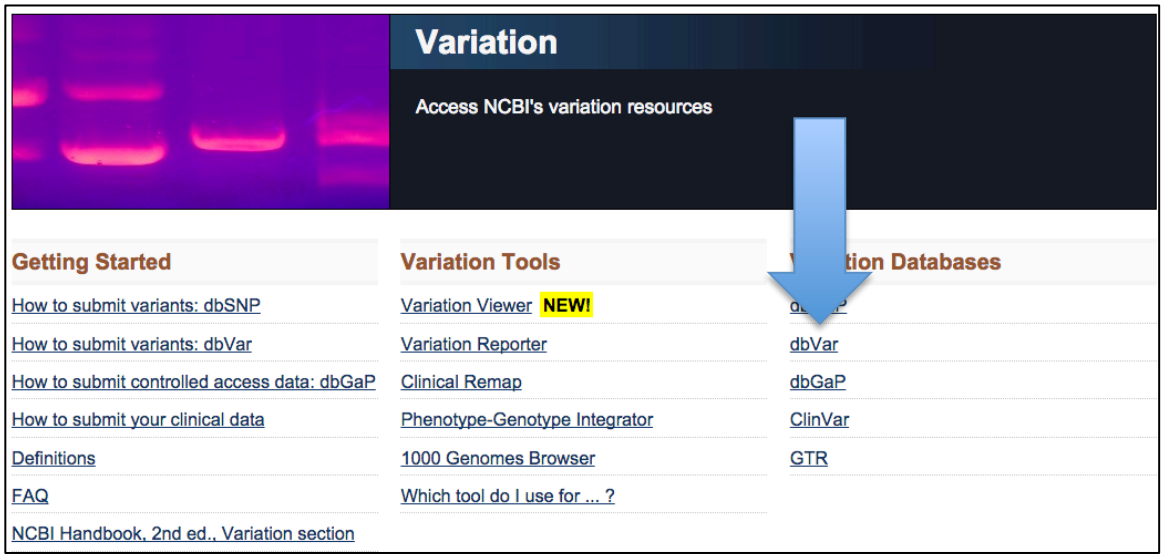
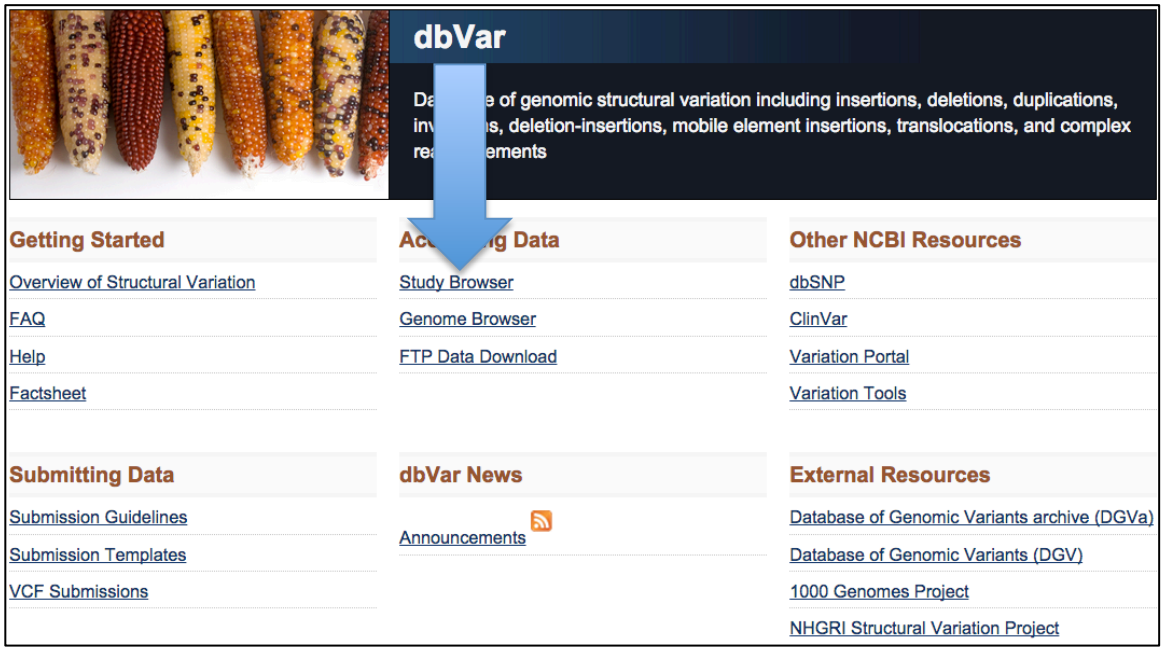
These demonstrations will lead you through the main databases and tools for genetic variation at the NCBI.

dbVAR

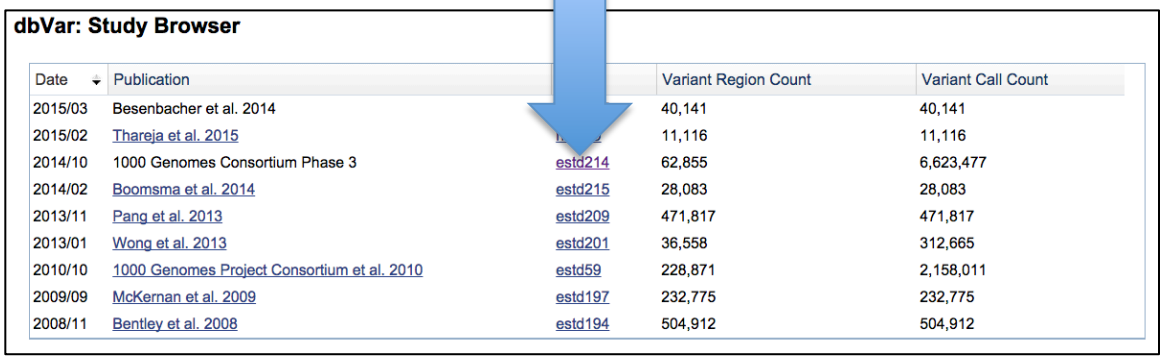
There are phase II drug metabolism enzyme genes that are missing in some individuals, notably some glutathione S-transferase genes (GSTM1, GSTM2 and GSTT2) and a UDP glucuronosyl transferase (UGT2B17). We can identify variant calls from the 1000 Genomes project data in dbVAR to see how prevalent these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



Click the link to the Study Browser



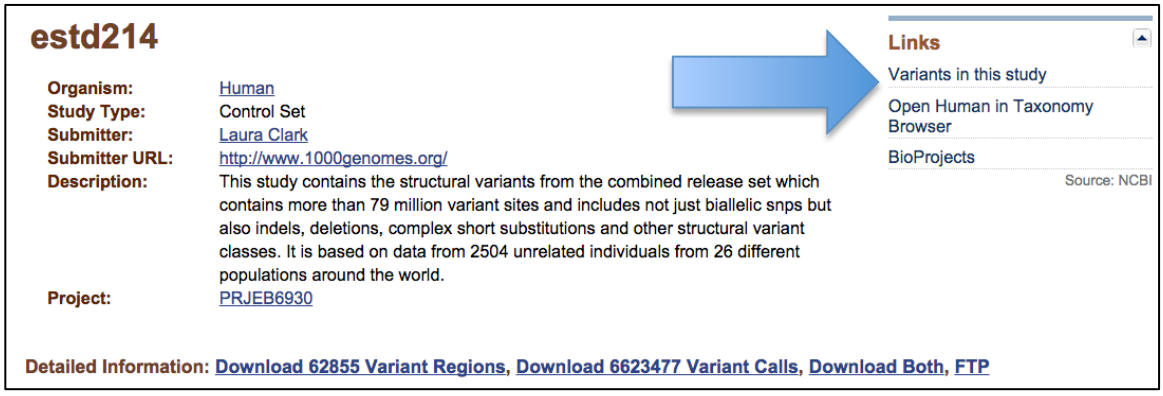
Use the Filters on the right-hand side of the browser to filter for sequencing studies of human with more than 10,000 variant regions.



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	estd214	62,855	6,623,477
2014/02	Boomsma et al. 2014	estd215	28,083	28,083
2013/11	Pang et al. 2013	estd209	471,817	471,817
2013/01	Wong et al. 2013	estd201	36,558	312,665
2010/10	1000 Genomes Project Consortium et al. 2010	estd59	228,871	2,158,011
2009/09	McKernan et al. 2009	estd197	232,775	232,775
2008/11	Bentley et al. 2008	estd194	504,912	504,912

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



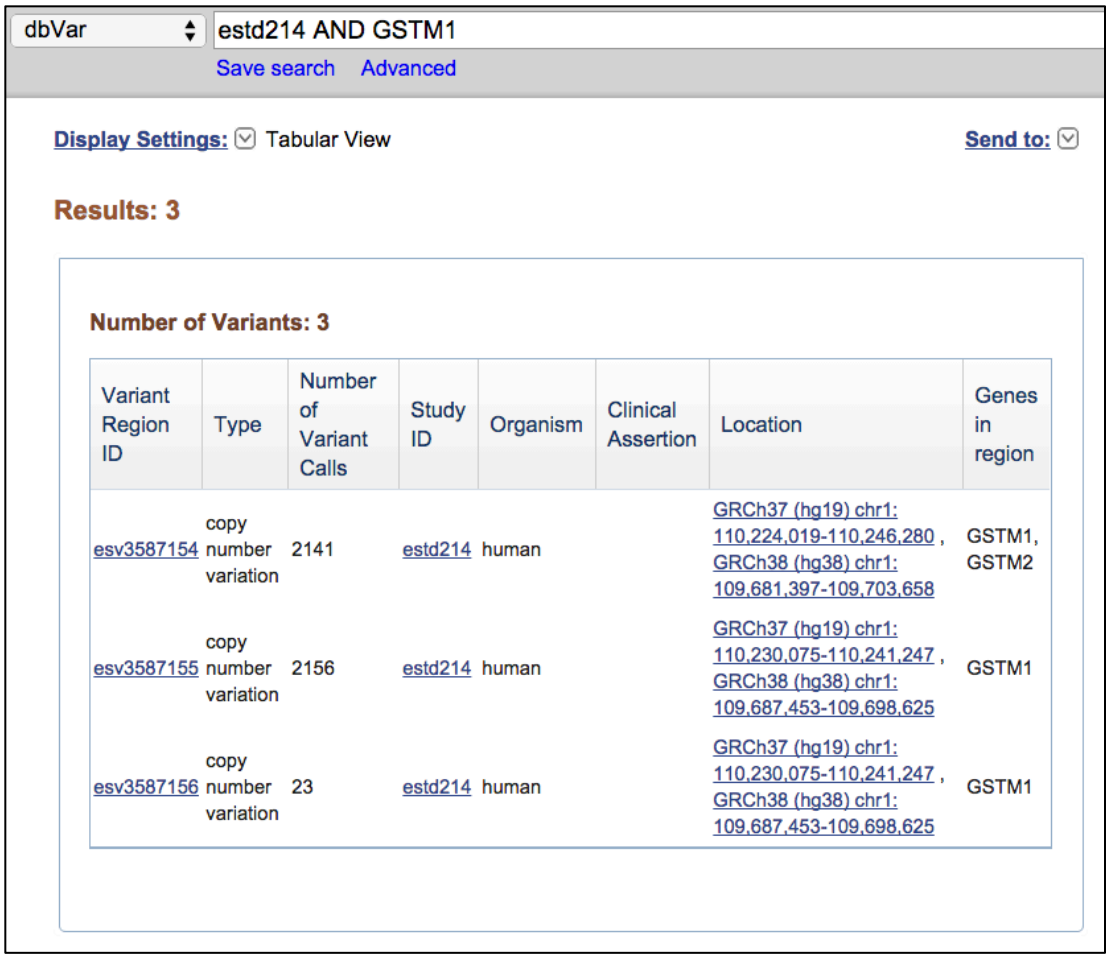
estd214

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

Links
Variants in this study
Open Human in Taxonomy Browser
BioProjects
Source: NCBI

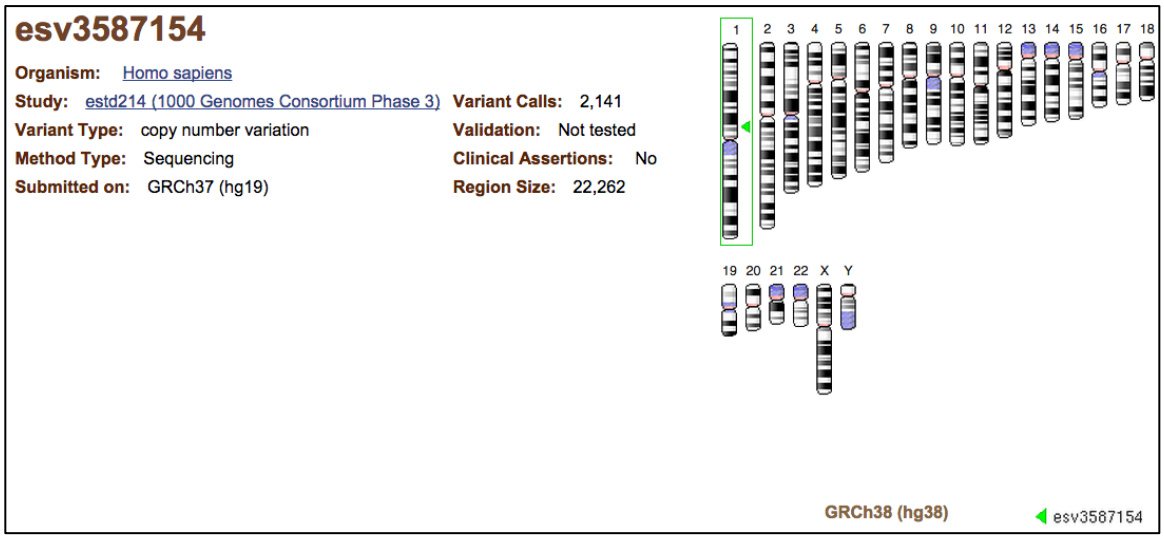
Detailed Information: Download 62855 Variant Regions, Download 6623477 Variant Calls, Download Both, FTP

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

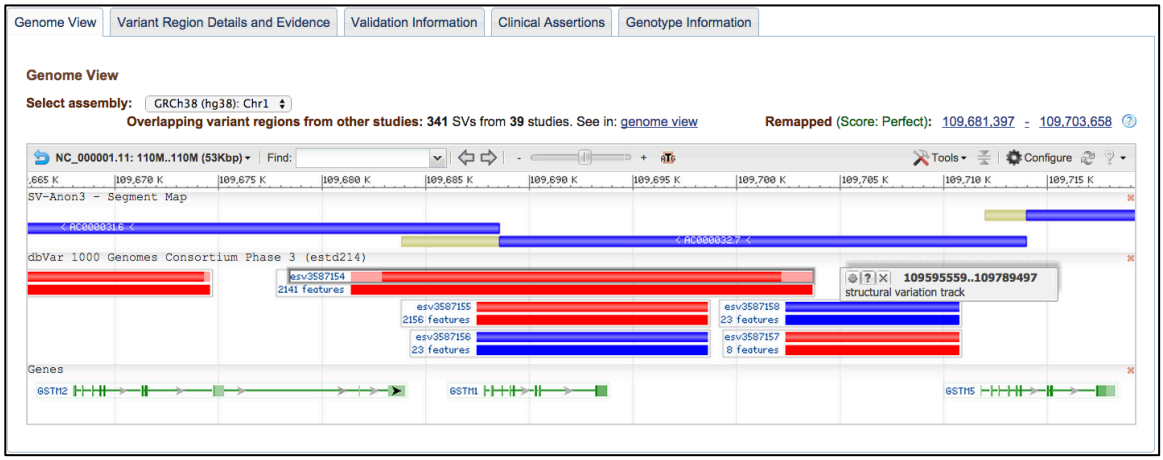


Variant Region ID	Type	Number of Variant Calls	Study ID	Organism	Clinical Assertion	Location	Genes in region
esv3587154	copy number variation	2141	estd214	human		GRCh37 (hg19) chr1: 110,224,019-110,246,280 , GRCh38 (hg38) chr1: 109,681,397-109,703,658	GSTM1, GSTM2
esv3587155	copy number variation	2156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



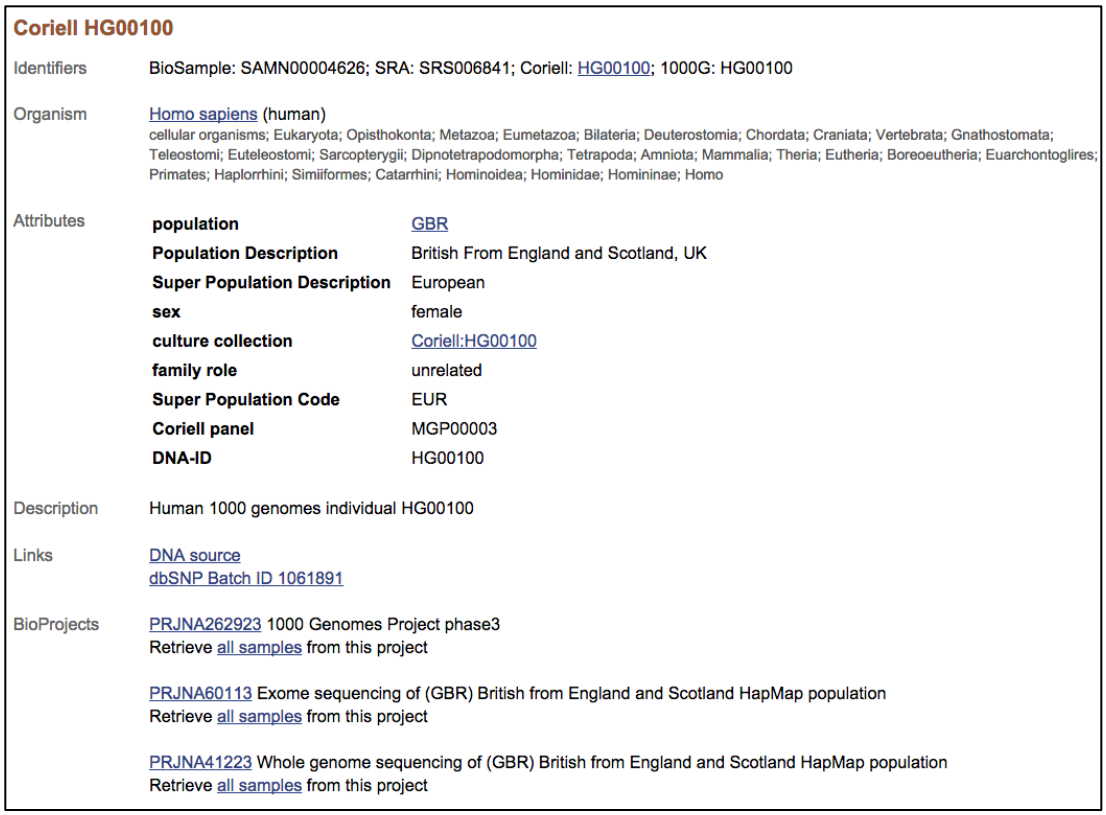
You can use the Configure menu to set the display options for the structural variation track to “Show parent, Merge children”, making the display simpler. Use the zoom feature on the viewer to expand to include both GSTM1 and GSTM2 genes.



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	SAMN00004622	Sequencing	Read depth and paired-end mapping	Homozygous	2,029
essv10056818	copy number loss	SAMN00004623	Sequencing	Read depth and paired-end mapping	Homozygous	2,569
essv10056819	copy number loss	SAMN00004625	Sequencing	Read depth and paired-end mapping	Homozygous	2,494
essv10056820	copy number loss	SAMN00004626	Sequencing	Read depth and paired-end mapping	Heterozygous	2,840
essv10056821	copy number loss	SAMN00004627	Sequencing	Read depth and paired-end mapping	Homozygous	2,576
essv10056822	copy number loss	SAMN00004628	Sequencing	Read depth and paired-end mapping	Heterozygous	2,591
essv10056823	copy number loss	SAMN00004631	Sequencing	Read depth and paired-end mapping	Homozygous	2,605
essv10056824	copy number loss	SAMN00004632	Sequencing	Read depth and paired-end mapping	Homozygous	2,578
essv10056825	copy number loss	SAMN00004633	Sequencing	Read depth and paired-end mapping	Homozygous	2,544
essv10056826	copy number loss	SAMN00004634	Sequencing	Read depth and paired-end mapping	Homozygous	2,579
essv10056827	copy number loss	SAMN00004635	Sequencing	Read depth and paired-end mapping	Homozygous	2,666
essv10056828	copy number loss	SAMN00004636	Sequencing	Read depth and paired-end mapping	Heterozygous	2,583
essv10056829	copy number loss	SAMN00004637	Sequencing	Read depth and paired-end mapping	Heterozygous	2,499
essv10056830	copy number loss	SAMN00004638	Sequencing	Read depth and paired-end mapping	Homozygous	2,269
essv10056831	copy number loss	SAMN00004639	Sequencing	Read depth and paired-end mapping	Heterozygous	2,601
essv10056832	copy number loss	SAMN00004640	Sequencing	Read depth and paired-end mapping	Homozygous	2,132
essv10056833	copy number loss	SAMN00004641	Sequencing	Read depth and paired-end mapping	Heterozygous	2,811
essv10056834	copy number loss	SAMN00004642	Sequencing	Read depth and paired-end mapping	Homozygous	2,561
essv10056835	copy number loss	SAMN00004643	Sequencing	Read depth and paired-end mapping	Homozygous	2,449
essv10056836	copy number loss	SAMN00006337	Sequencing	Read depth and paired-end mapping	Heterozygous	2,786
essv10056837	copy number loss	SAMN00006338	Sequencing	Read depth and paired-end mapping	Heterozygous	2,630
essv10056838	copy number loss	SAMN00004644	Sequencing	Read depth and paired-end mapping	Heterozygous	2,600
essv10056839	copy number loss	SAMN00006339	Sequencing	Read depth and paired-end mapping	Homozygous	2,641
essv10056840	copy number loss	SAMN00006340	Sequencing	Read depth and paired-end mapping	Heterozygous	2,632

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

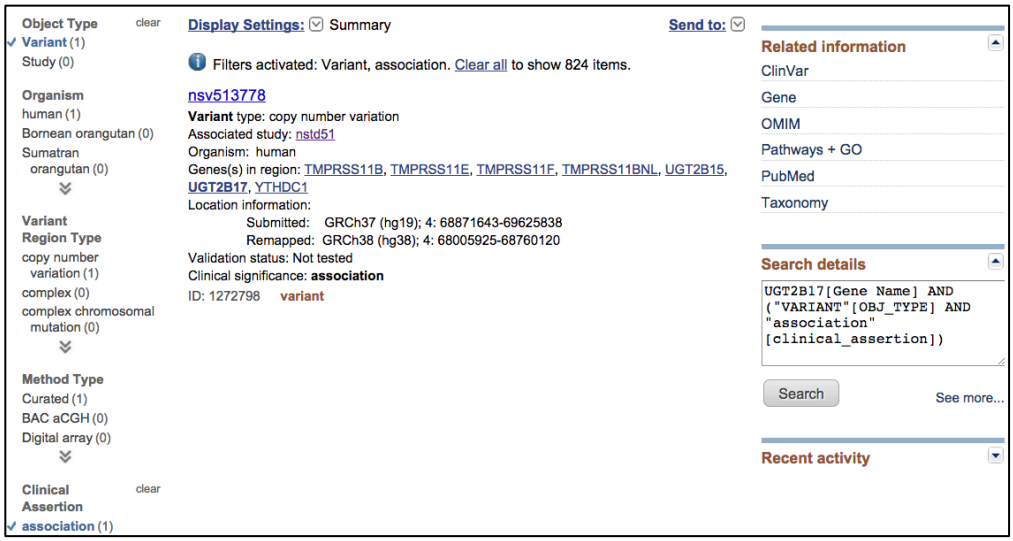


Coriell HG00100

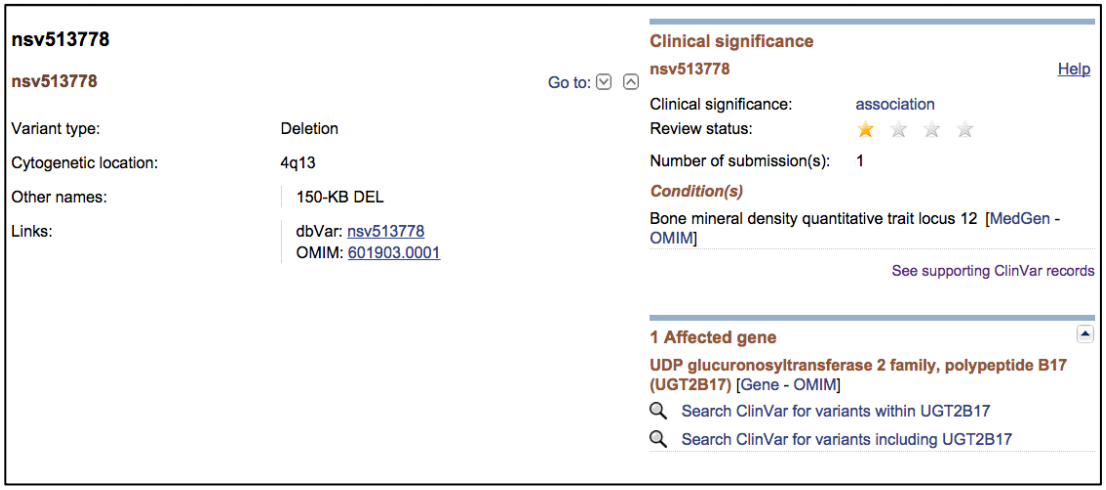
Identifiers	BioSample: SAMN00004626; SRA: SRS006841; Coriell: HG00100; 1000G: HG00100
Organism	Homo sapiens (human) cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo
Attributes	**population**: GBR **Population Description**: British From England and Scotland, UK **Super Population Description**: European **sex**: female **culture collection**: Coriell:HG00100 **family role**: unrelated **Super Population Code**: EUR **Coriell panel**: MGP00003 **DNA-ID**: 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

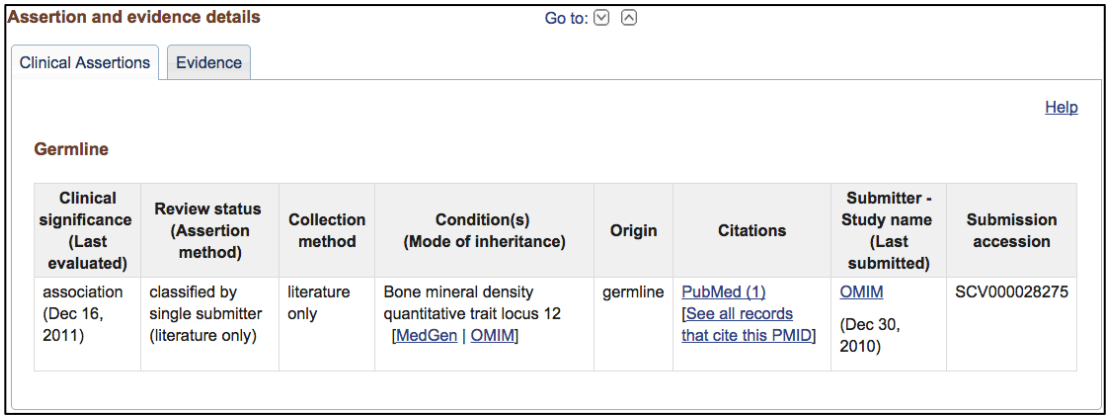
There are only 100 variant calls in the table. You could get the complete set (VCF) from the 1000 genomes FTP directory. Notice that you can load all structural variants in the region into the dbVAR Genome Browser. This is an application that is very similar to the Variation Viewer that we will use later.

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".



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





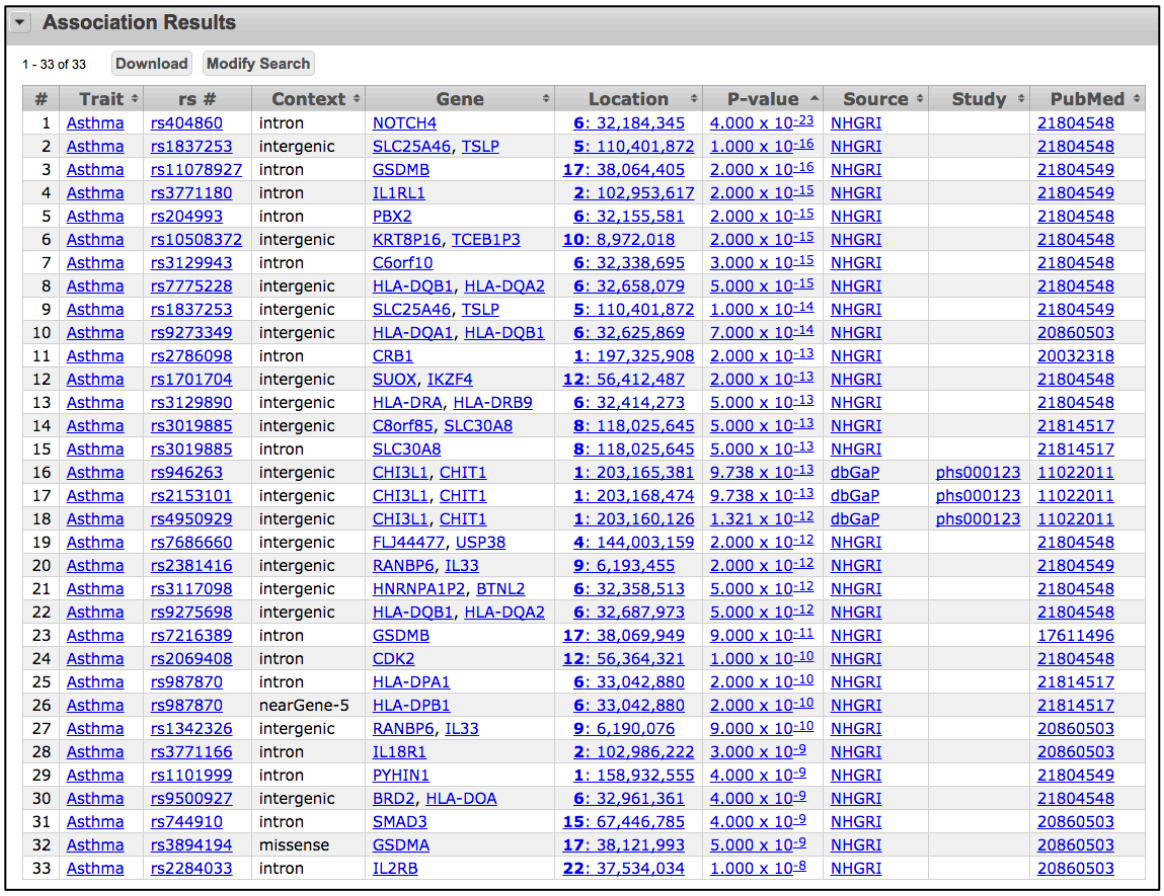
Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter - Study name (Last submitted)	Submission accession
association (Dec 16, 2011)	classified by single submitter (literature only)	literature only	Bone mineral density quantitative trait locus 12 [MedGen | OMIM]	germline	PubMed (1) [See all records that cite this PMID]	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 1×10^{-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.

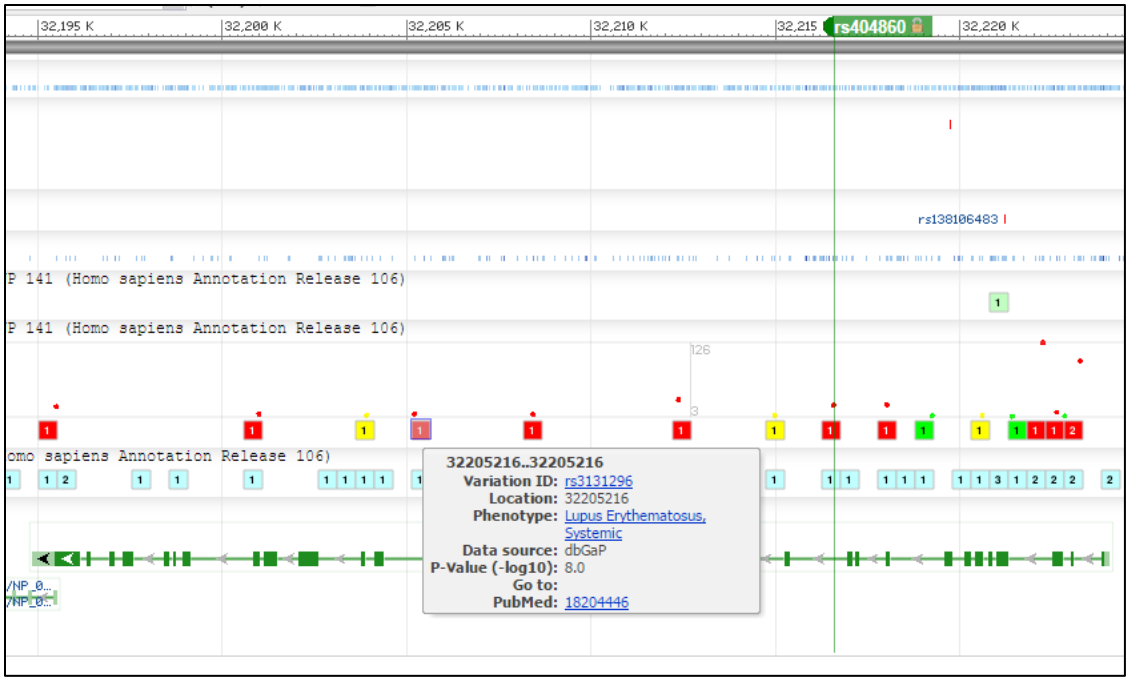


Association Results

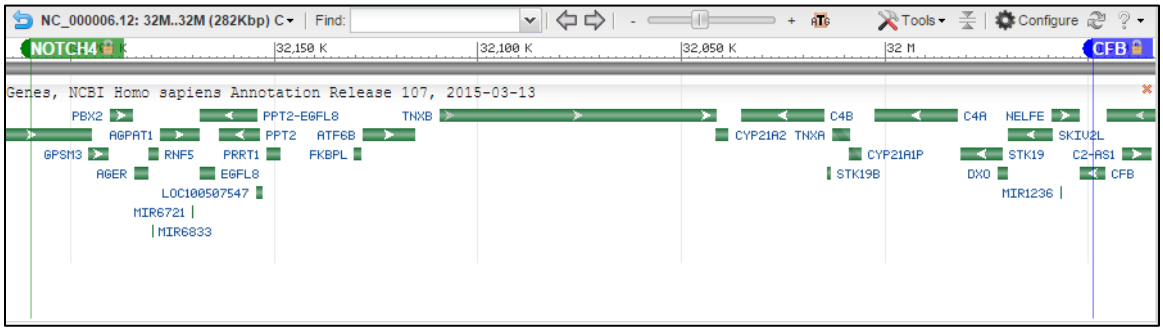
1 - 33 of 33 Download Modify Search

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



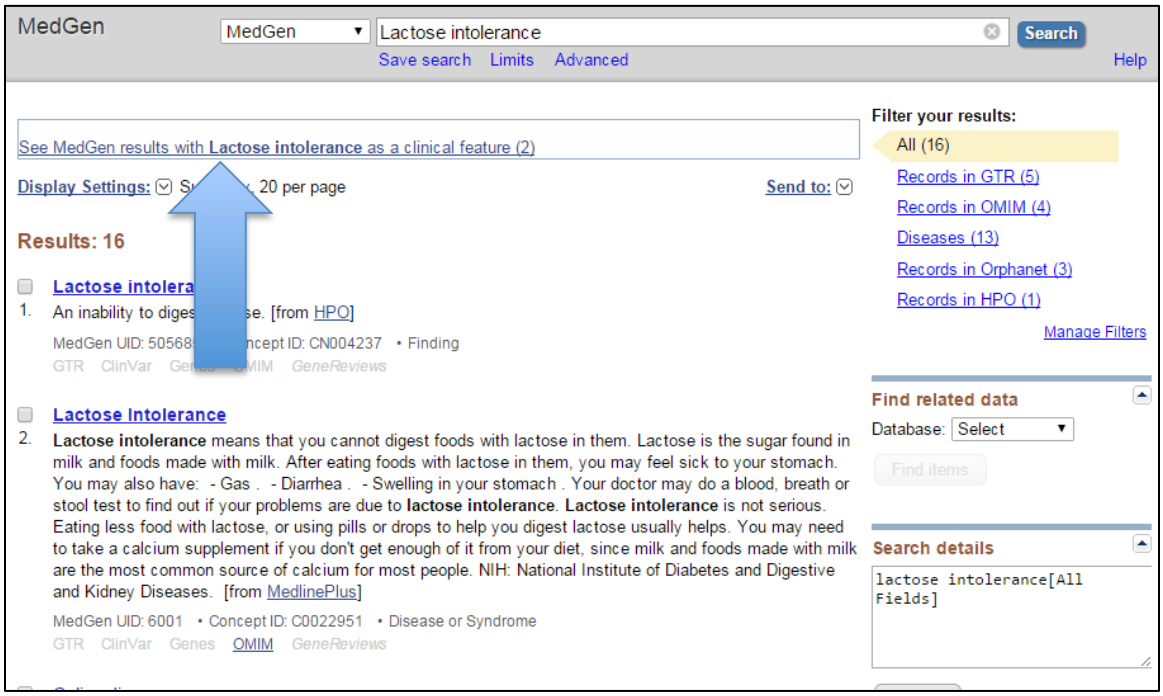
Another section of the report shows the dbGaP studies that have asthma as a measured variable. You would need to apply to get access to the individual level data.

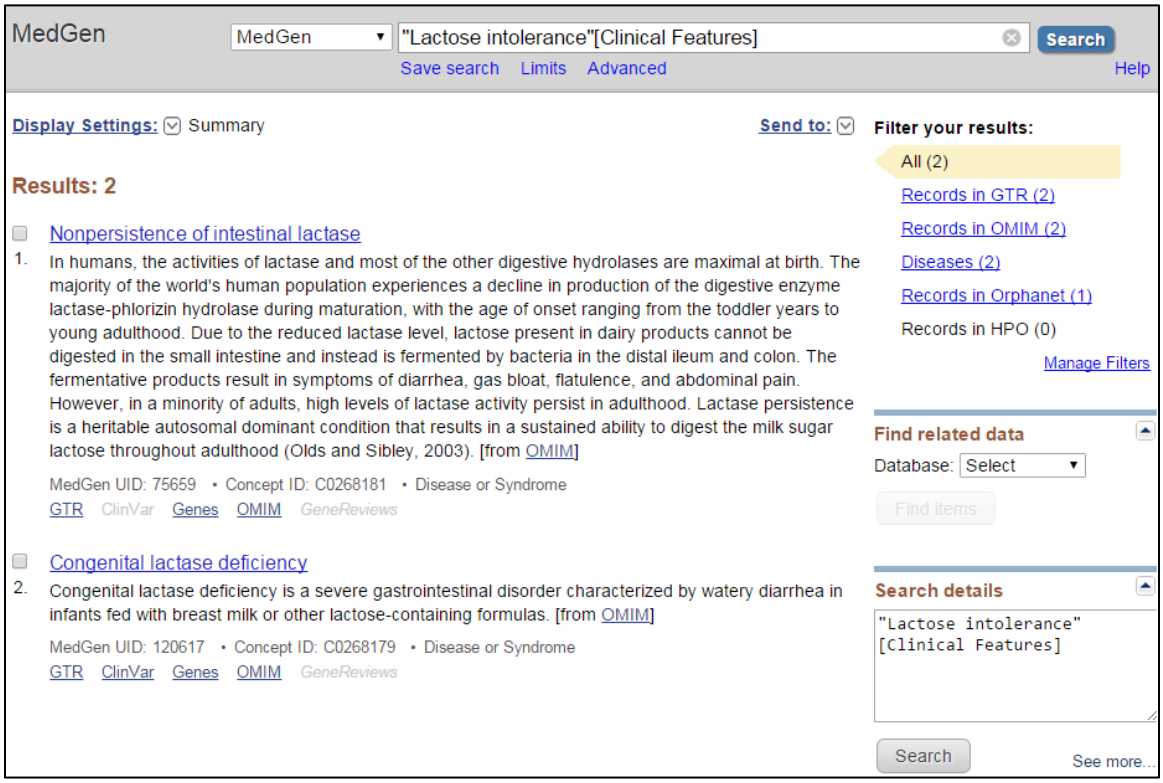
Independent exercise: finding association results for serum triglycerides

- 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?

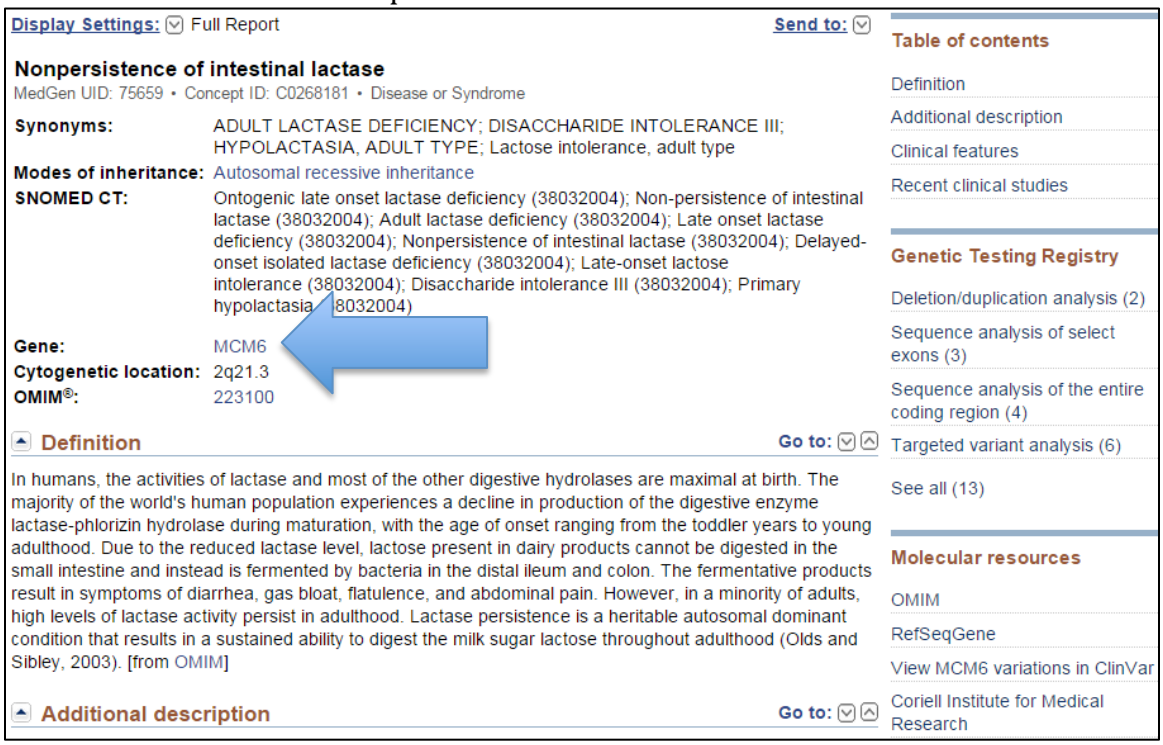
Lactase persistence in MedGen, ClinVar, 1000Genomes

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

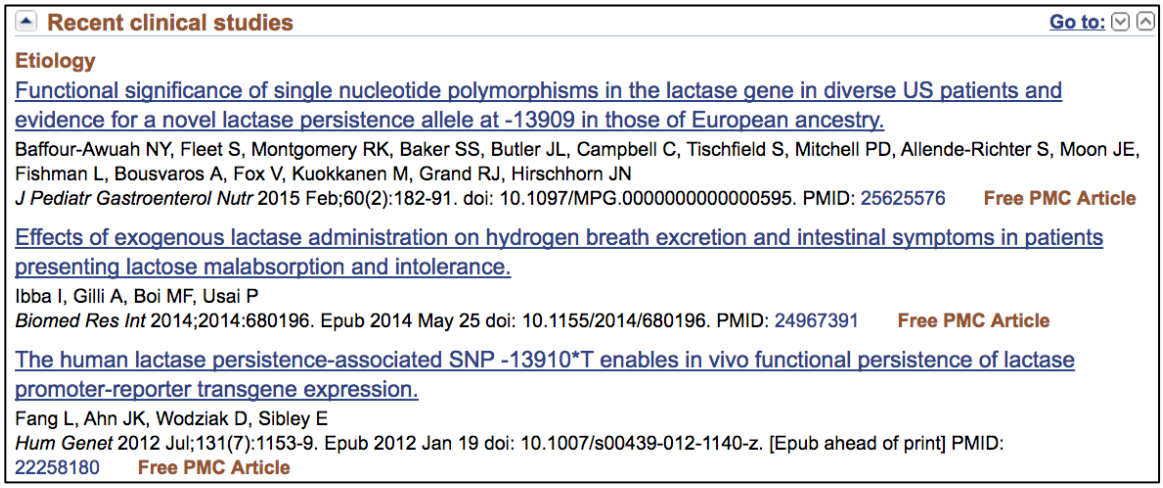




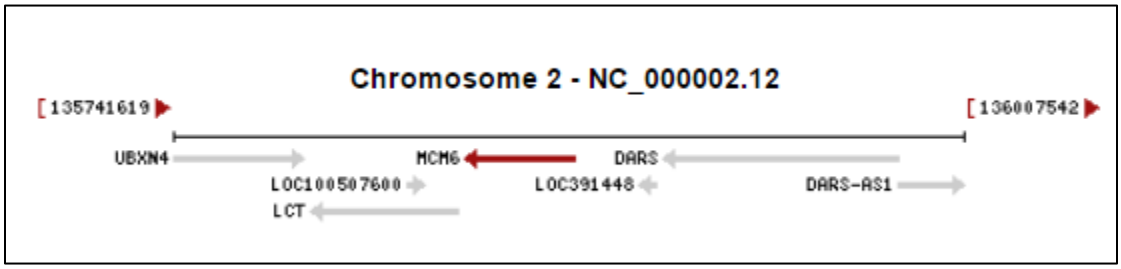
Retrieve the record for "Nonpersistence of intestinal lactase".



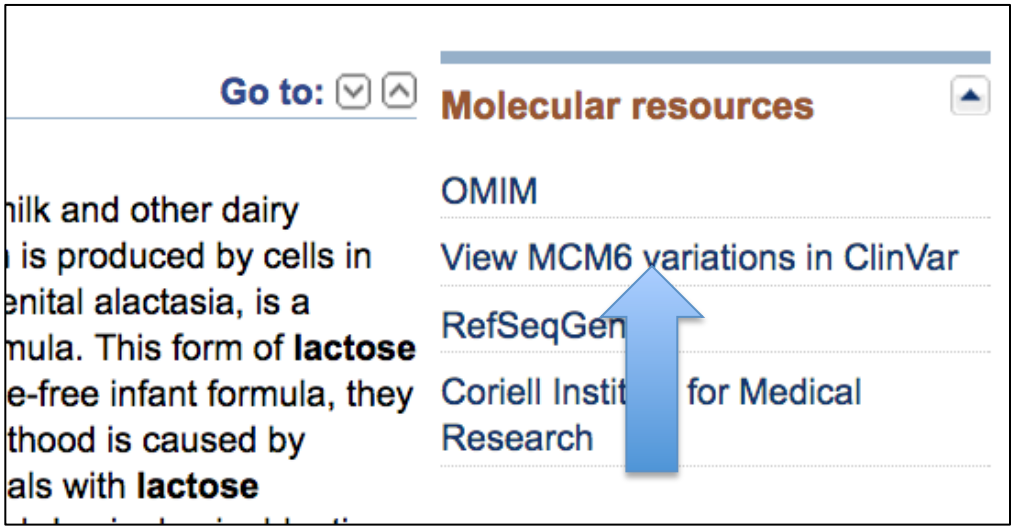
MedGen is a largely automatic aggregator of information on phenotypes. It includes several controlled vocabularies for diseases / disorders including SNOMED CT, MeSH, and the vocabulary used by GTR. Notice that this condition is labeled with the MCM6 gene rather than the lactase gene. If you look at 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.



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.

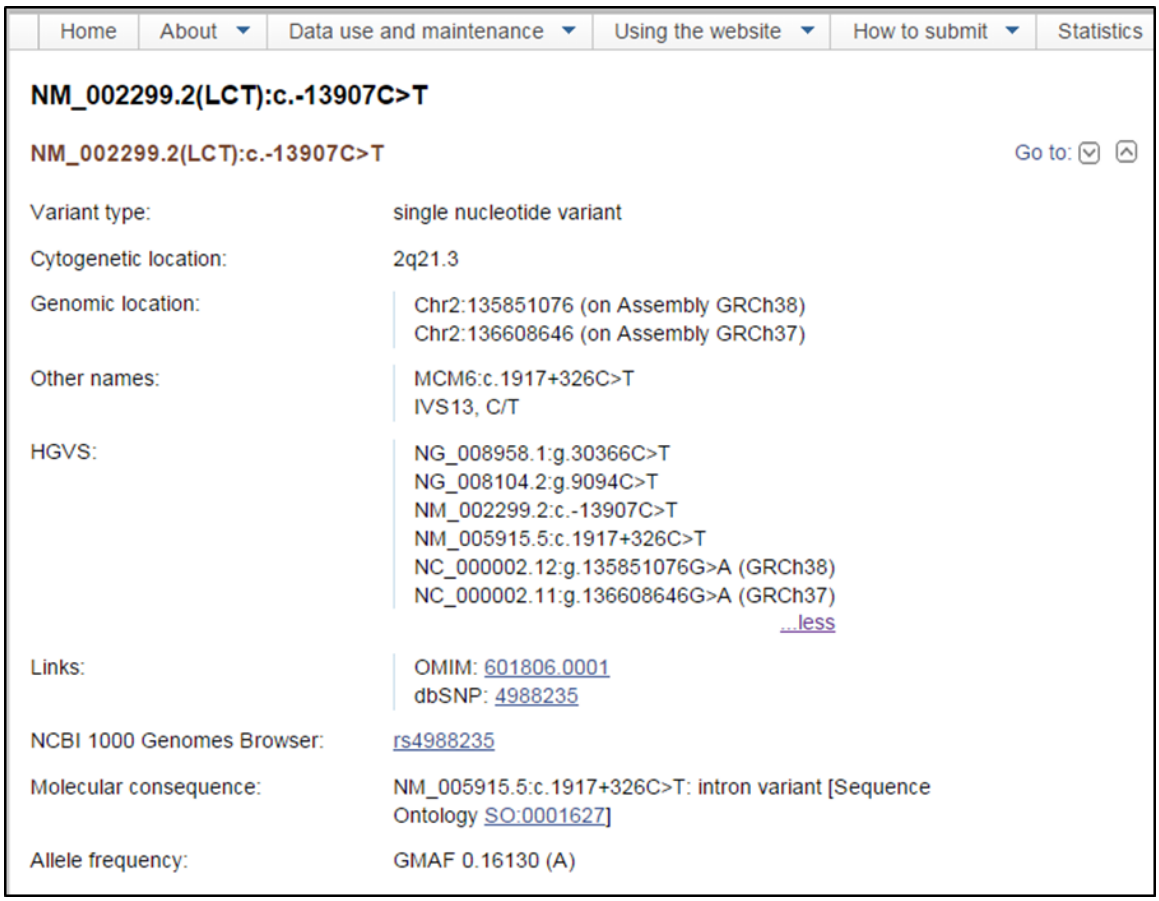


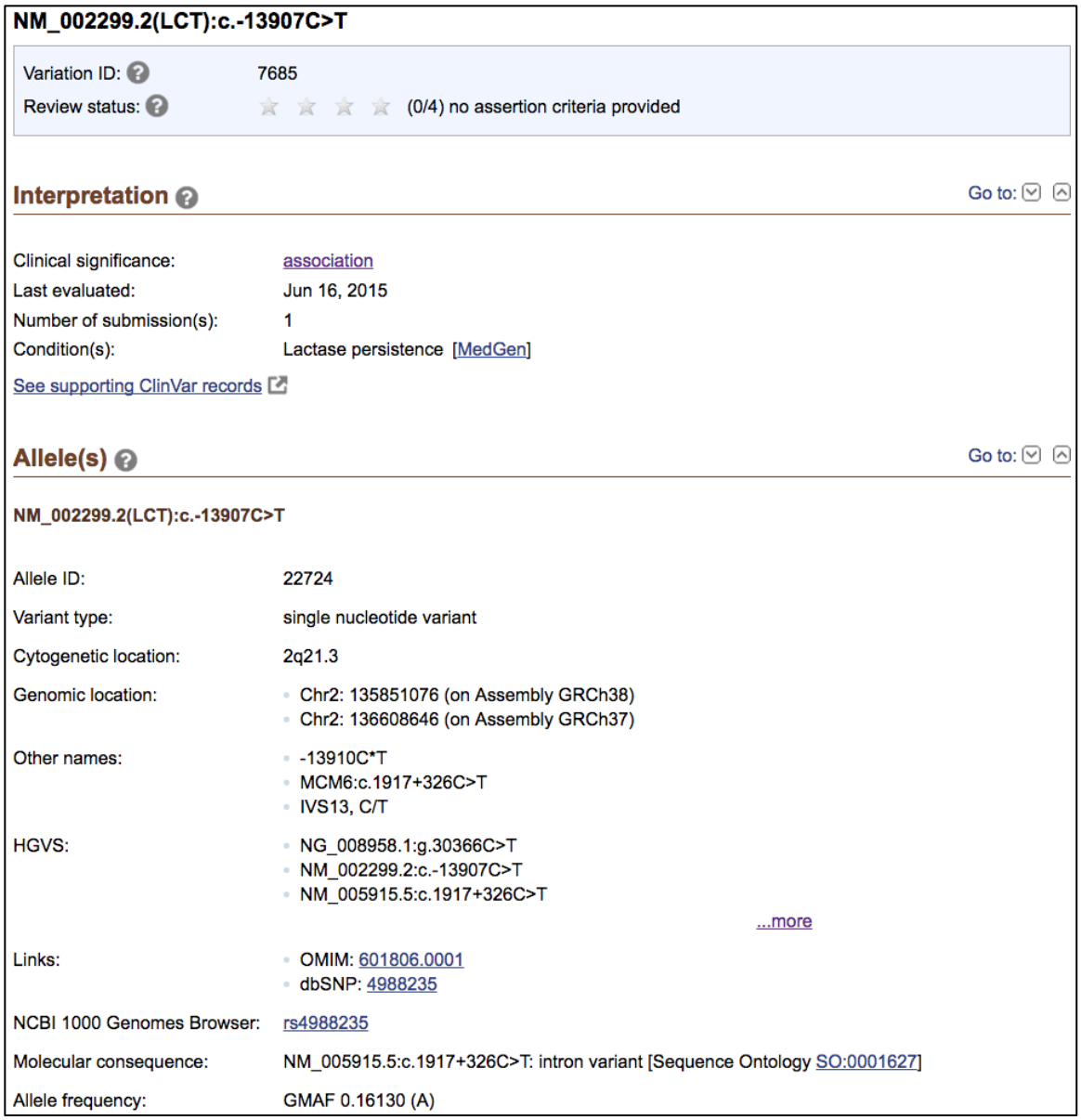
Go back to the MedGen record and click the link to ClinVar under Molecular Resources.



	Variation *Location*	Gene(s)	Condition(s)	Frequency	Clinical significance (Last reviewed)
1.	MCM6, EX17, T/C	MCM6	Lactate persistence		Pathogenic (Jan 1, 2008)
2.	NM_005915.5(MCM6):c.1917+329C>G *GRCh37:* Chr2:136608643 *GRCh38:* Chr2:135851073	MCM6	Lactase persistence		association (Jun 16, 2015)
3.	NM_002299.2(LCT):c.-13907C>T *GRCh37:* Chr2:136608646 *GRCh38:* Chr2:135851076	MCM6	Lactase persistence	GMAF:0.16130(A)	association (Jun 16, 2015)
4.	NM_005915.5(MCM6):c.1917+321T>G *GRCh37:* Chr2:136608651 *GRCh38:* Chr2:135851081	MCM6	Lactase persistence	GMAF:0.00060(C)	association (Jun 16, 2015)
5.	NM_005915.5(MCM6):c.1917+226G>C *GRCh37:* Chr2:136608746 *GRCh38:* Chr2:135851176	MCM6	Lactase persistence	GMAF:0.00340(G)	association (Jun 16, 2015)
6.	NM_005915.5(MCM6):c.1362+117G>A *GRCh37:* Chr2:136616754 *GRCh38:* Chr2:135859184	MCM6	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:

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:

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

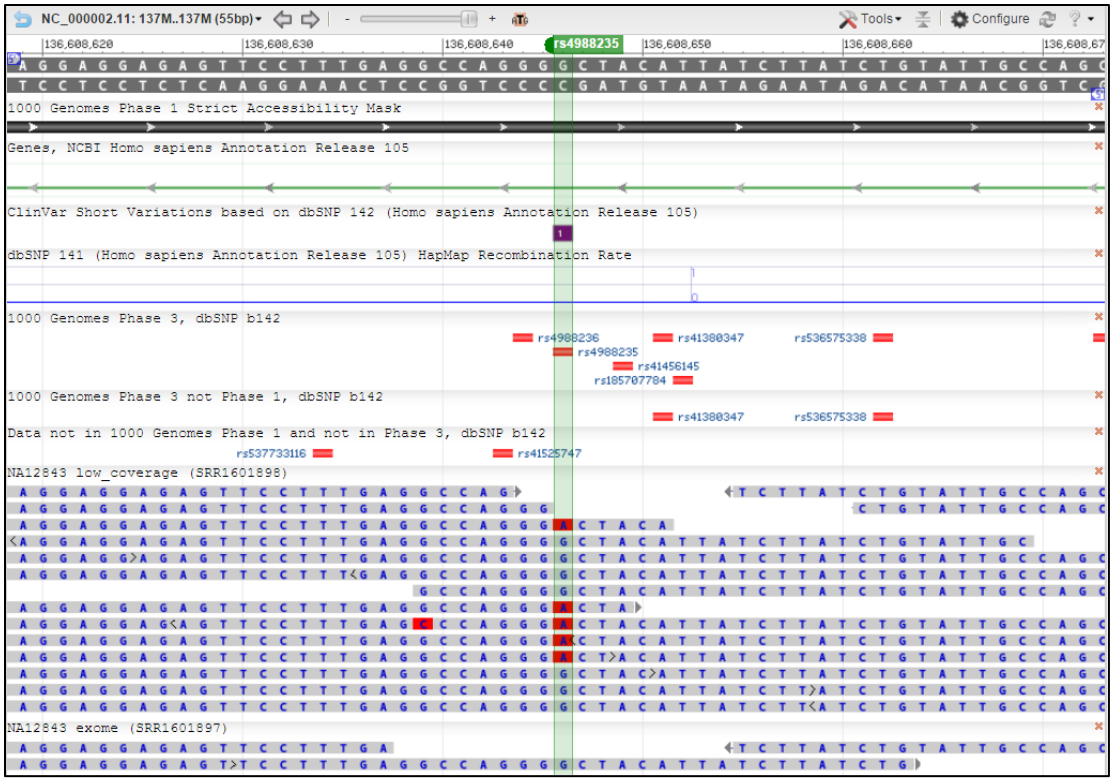
NCBI 1000 Genomes Browser: rs4988235
Molecular consequence: NM_005915.5:c.1917+326C>T: intron variant [Sequence Ontology SO:0001627]
Allele frequency: GMAF 0.16130 (A)

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

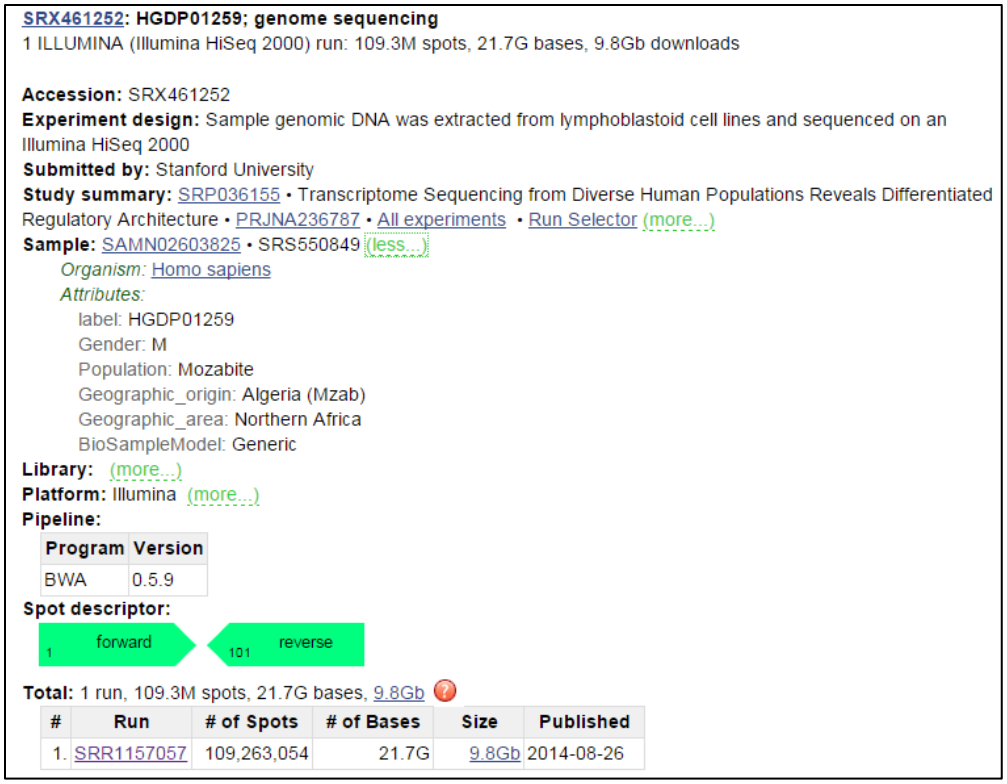
actually is the major allele in populations from northern Europe – greater than 70% in the Utah European ancestry (CEU) and British populations (GBR), 59% in Finnish (Fin), but absent from the Asian populations. The persistence allele is thought to have increased in frequency because of selection pressure 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
Populations / Samples Show: Allele frequencies	**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 Carribbeans ...	**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 Bejin...	**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:

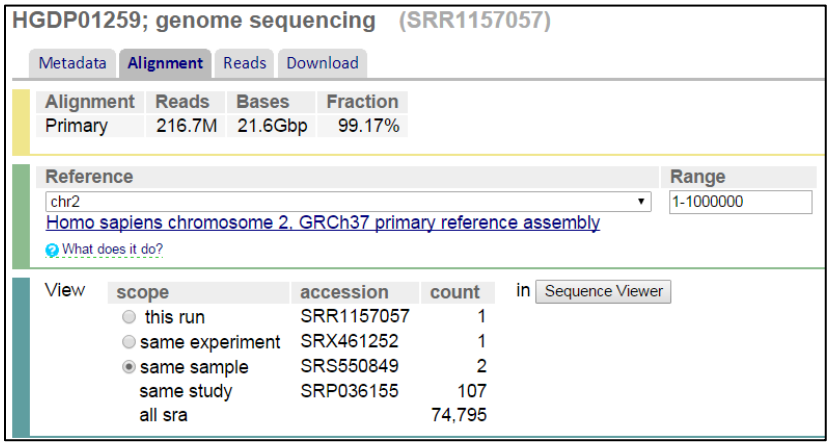
Program	Version
BWA	0.5.9

Spot descriptor:
1 forward 101 reverse

Total: 1 run, 109.3M spots, 21.7G bases, 9.8Gb

#	Run	# of Spots	# of Bases	Size	Published
1.	SRR1157057	109,263,054	21.7G	9.8Gb	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%

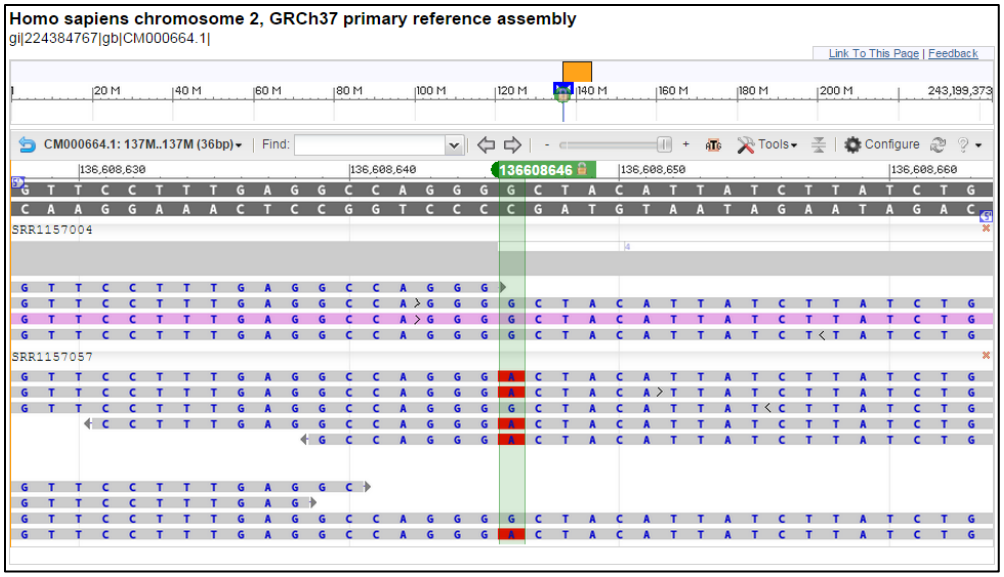
Reference: chr2 — Homo sapiens chromosome 2, GRCh37 primary reference assembly
What does it do?

Range: 1-1000000

View in Sequence Viewer

scope	accession	count
this run	SRR1157057	1
same experiment	SRX461252	1
same sample	SRS550849	2
same study	SRP036155	107
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.



This person is a probably heterozygote for the lactase persistence allele. This is consistent with adaptation an ancestral pastoralist lifestyle that included consumption of dairy.

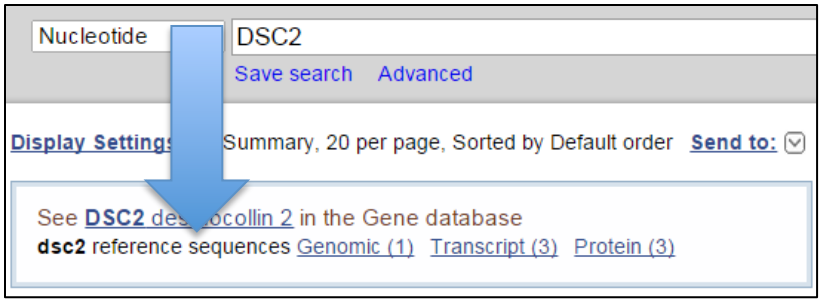
Go back and follow the link to dbSNP. You could use the FASTA sequence present in the SNP record to BLAST against the J. Craig Venter assembly (HuRef) to check which allele is included in his assembly.

Independent exercise: eye color variants

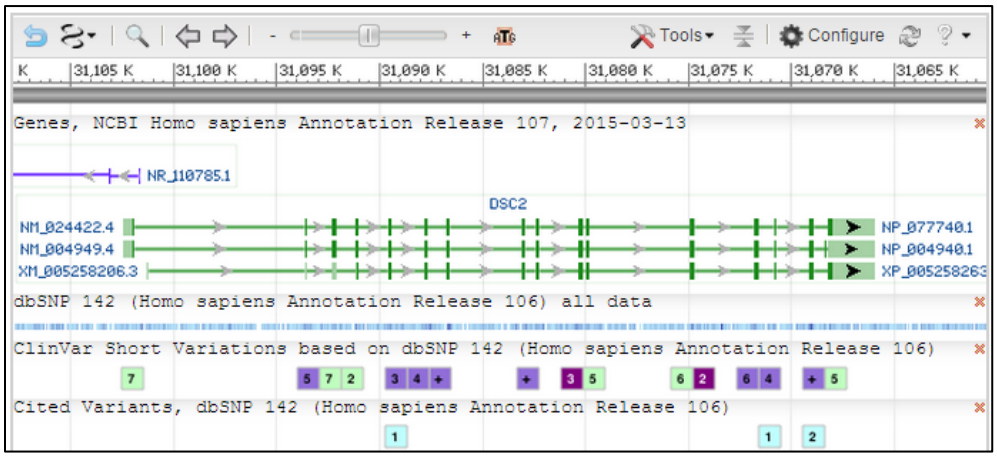
- Search MedGen for “blue eye color”.
- Retrieve the record for “Skin/hair/eye pigmentation, variation in, 1”.
- Follow the related information link to ClinVar
- Investigate the variant with the minor allele frequency of 18%. What populations are most likely to have the same allele as the Reference Genome? In which populations is the global minor allele now the major allele?
- What color eyes would the individual NA06984 likely have?

Using the variation reporter / variation viewer with desmocolin genes

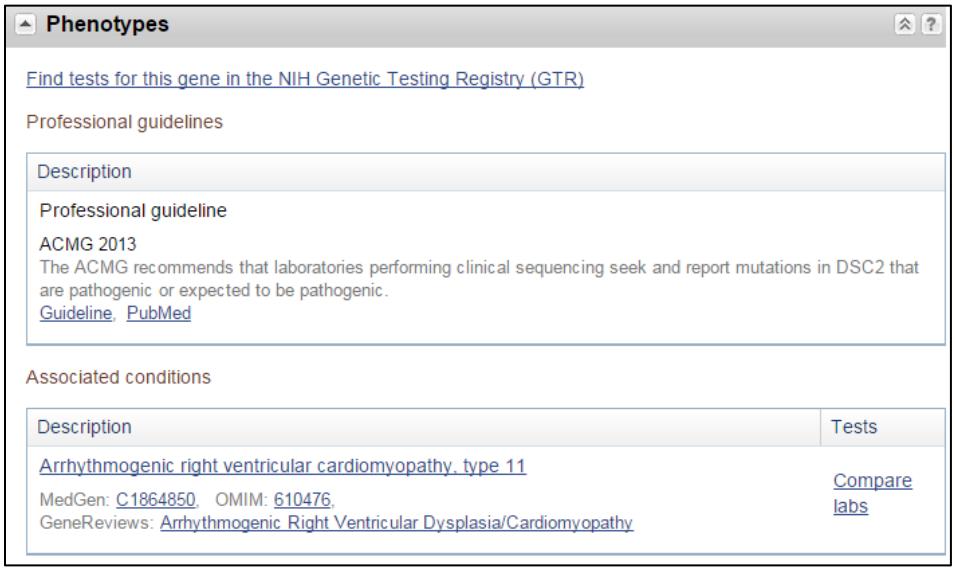
Use the Gene Sensor in the Nucleotide database to access the DSC2 gene record for human.



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



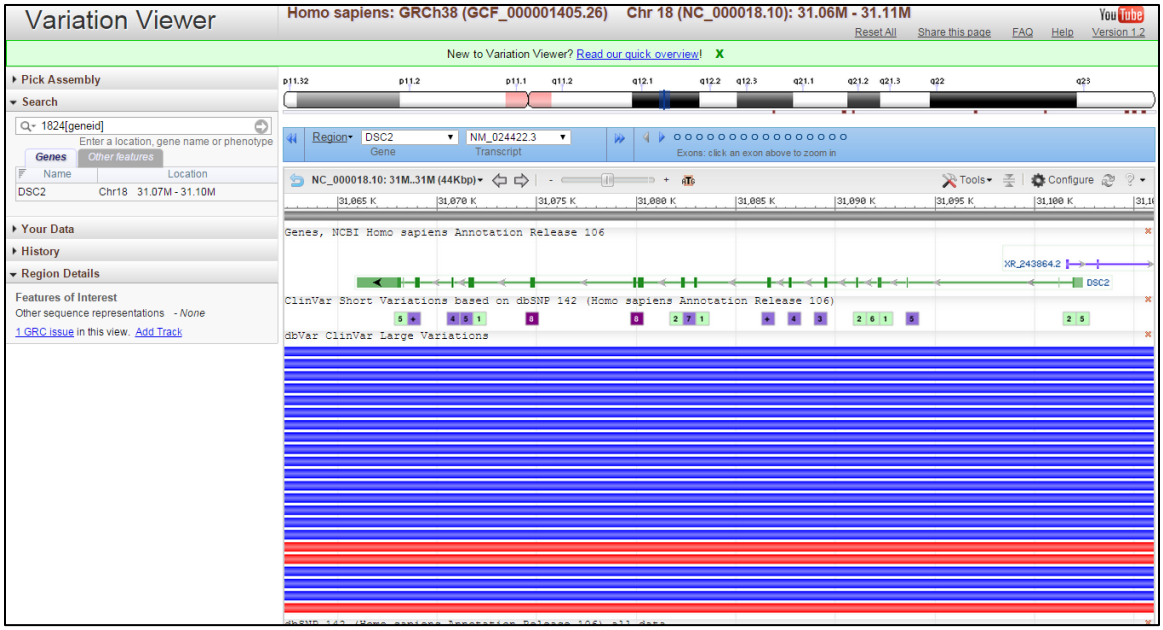
The Phenotypes section of the record suggests the dangerous nature of some of these as there is an ACMG guideline to report about the presence of these variants even as incidental findings.



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.

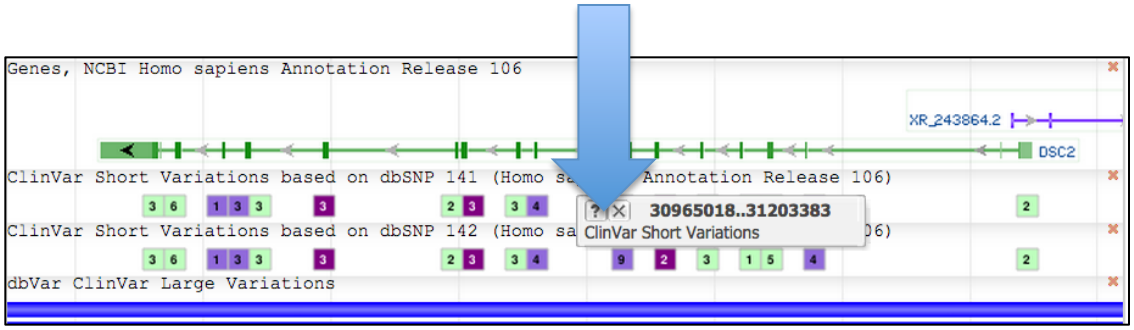
Accessing the Variation Viewer

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



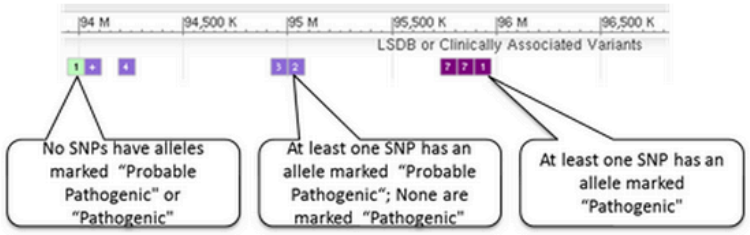
Tracks in the sequence viewer display

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.



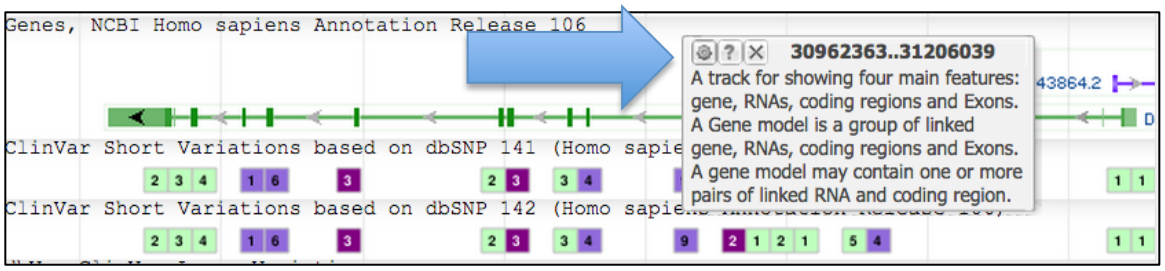
4.3. SNP Bins For Clinical Associations

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"

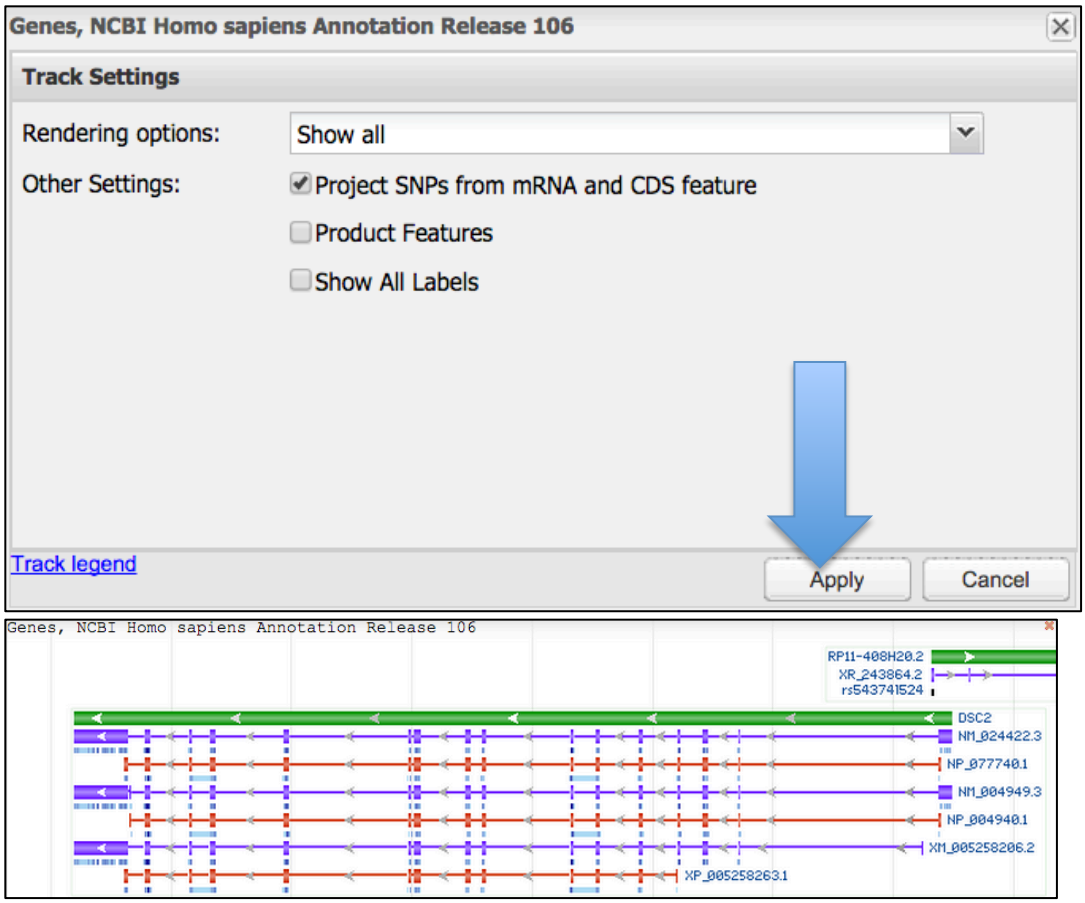


The next track shows large-scale variant regions from dbVar that have clinical assertions in ClinVar. These cover many genes. The red ones are relative deletions. The blue ones are duplications. You can check this with the "Track legend". The track at the bottom shows the locations of all variants from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

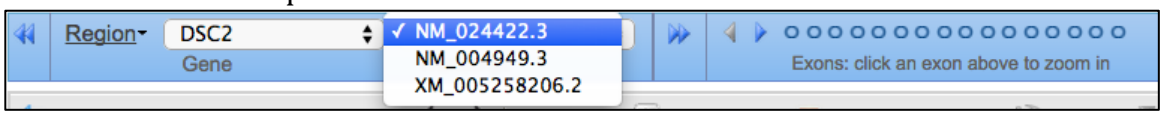
Change the Genes display to in the Variation Viewer to show all of the splice variants. Mouse over the top of the track and click the gear icon.



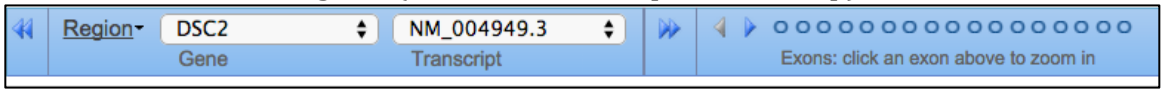
The Configure dialog appears for the Genes track. Set "Show all" in the rendering options and also check the "Project SNPs from mRNA and CDS feature", then click configure to apply these.



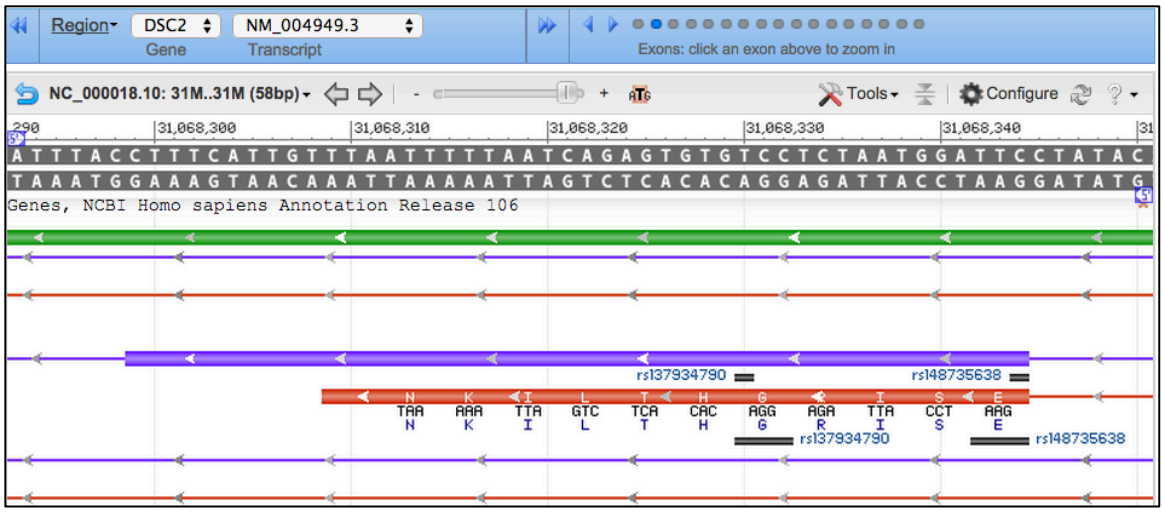
We can now see that there are three splice variants for this gene with slightly different sets of exons. We can use the Gene/Exon navigator to jump to the exon for the different transcript.



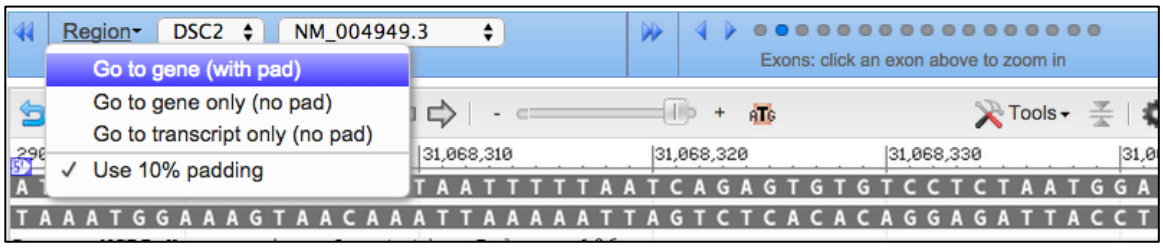
For example if I select NM_004949 as the transcript, notice that the number of exons shown in the exon navigator (the radio button spots at the top) increases.



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.

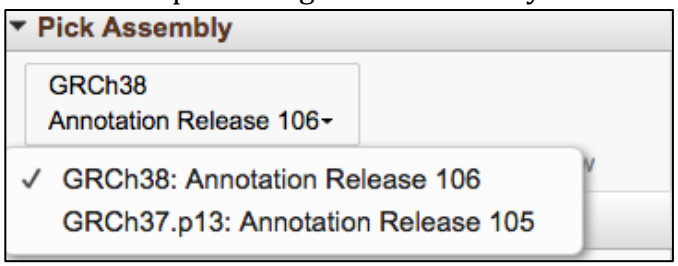


You are now zoomed in to the sequence level and can see nucleotides and amino acid sequences of the gene, transcript and products as well as individual SNP positions. You can restore the view using the Exon/Gene Navigator to select "Go to gene (with pad)" from the "Region" menu.



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.



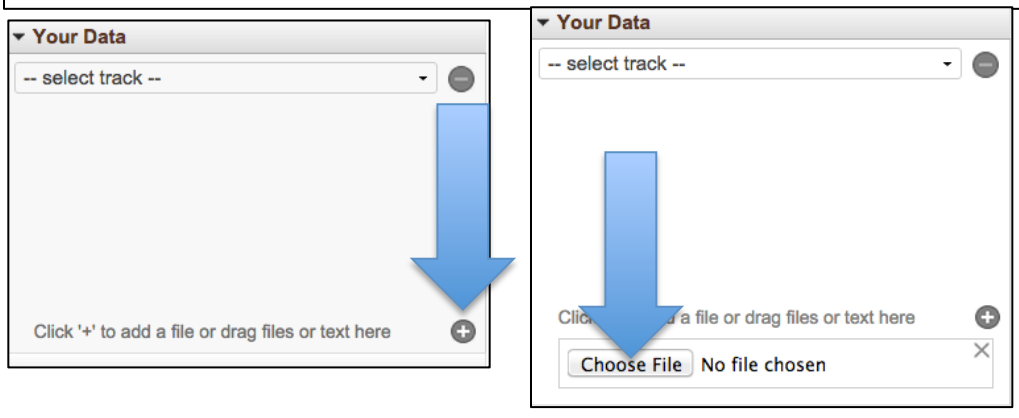
The "Region details" is where you may see other sequence representations such as alternate assemblies of the region representing other haplotypes if these were available. You may also see known issues with the current GRC build. In this case

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.

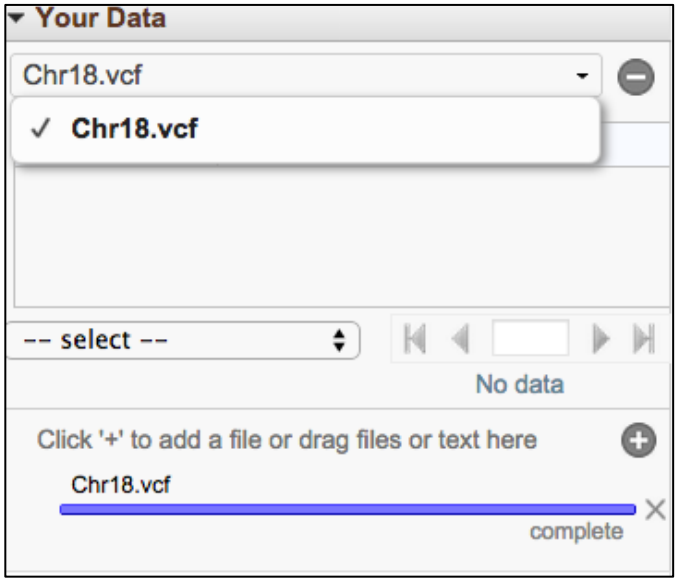
Uploading your own data

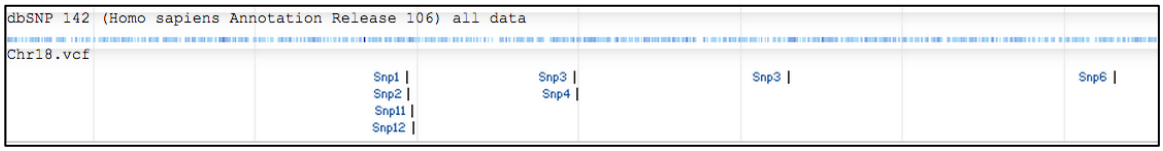
Before we leave this area of the viewer, notice that you can load your own data into the Variation Viewer. You'll need a file with your variant calls. The upload feature accepts a number of different formats including BED, HGVS, GVF and VCF. You can then use the "Your Data" section to upload your data. 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
```



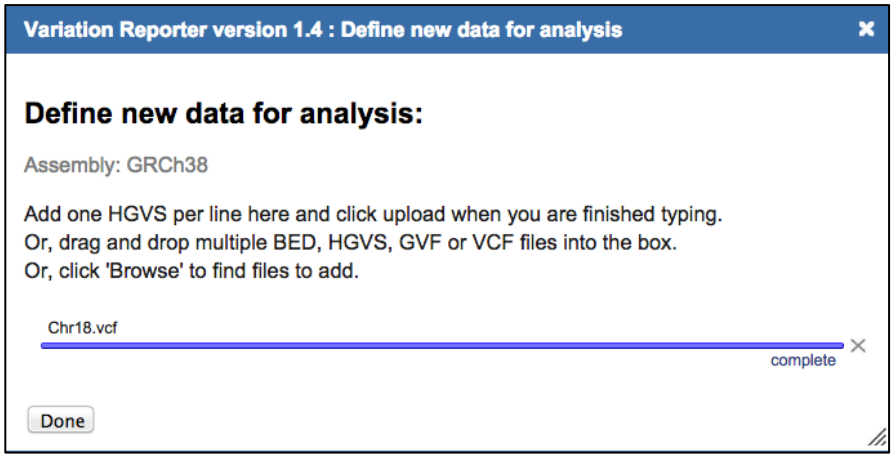
Select the track to add it to the display.

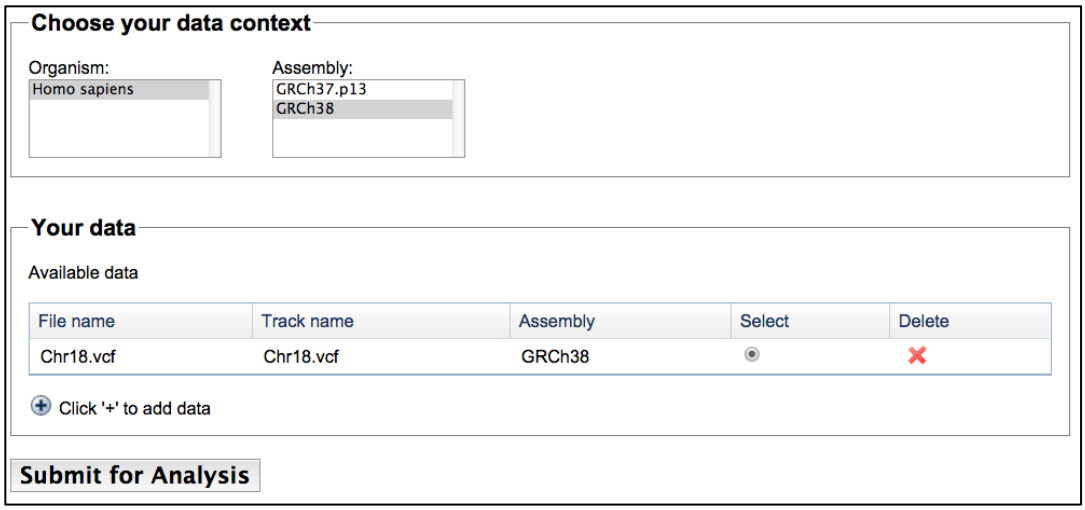


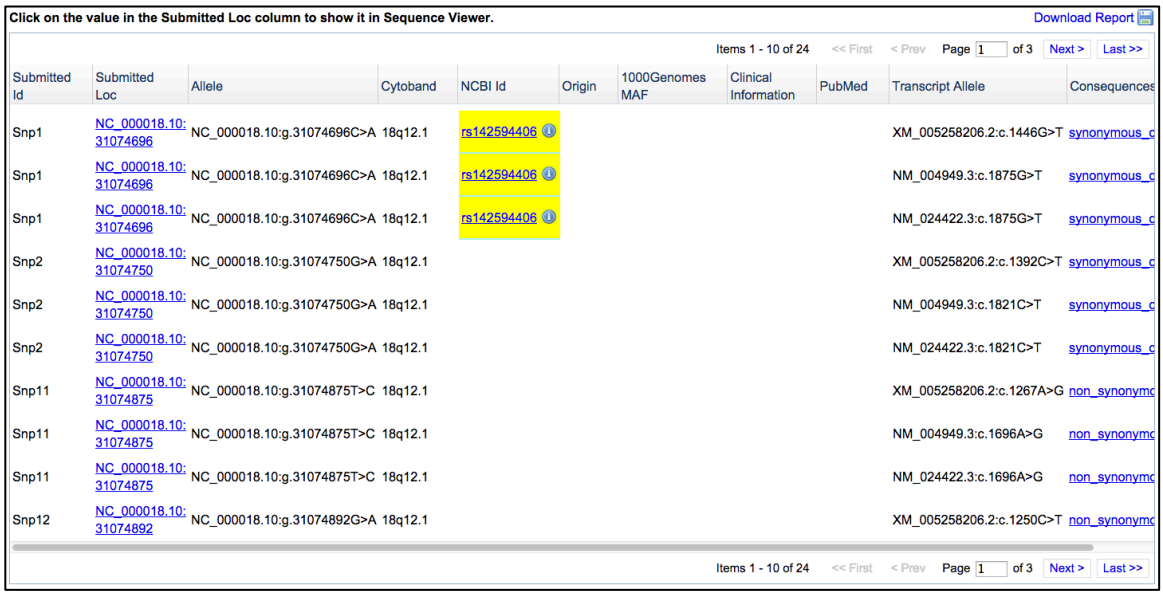


The Variation Reporter

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), which is linked to the Variation Portal page. Download the file to your computer and upload into the Variation Reporter.





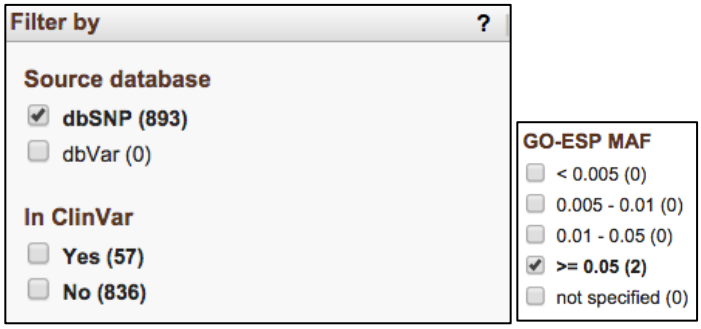


The Variation Reporter is particularly useful if you have your own variant calls and want to map them on to the genome and products. Notice that it provides the functional consequences for the variant calls and indicate when they match known SNPs. Follow the link to Snp11 to display it in the graphical sequence viewer at the bottom of the output

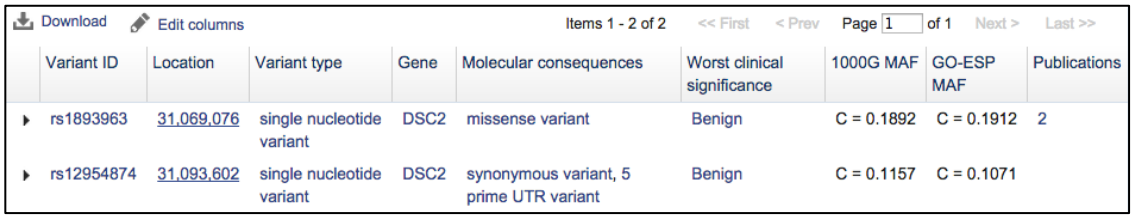
Using Filters and Variant Table

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

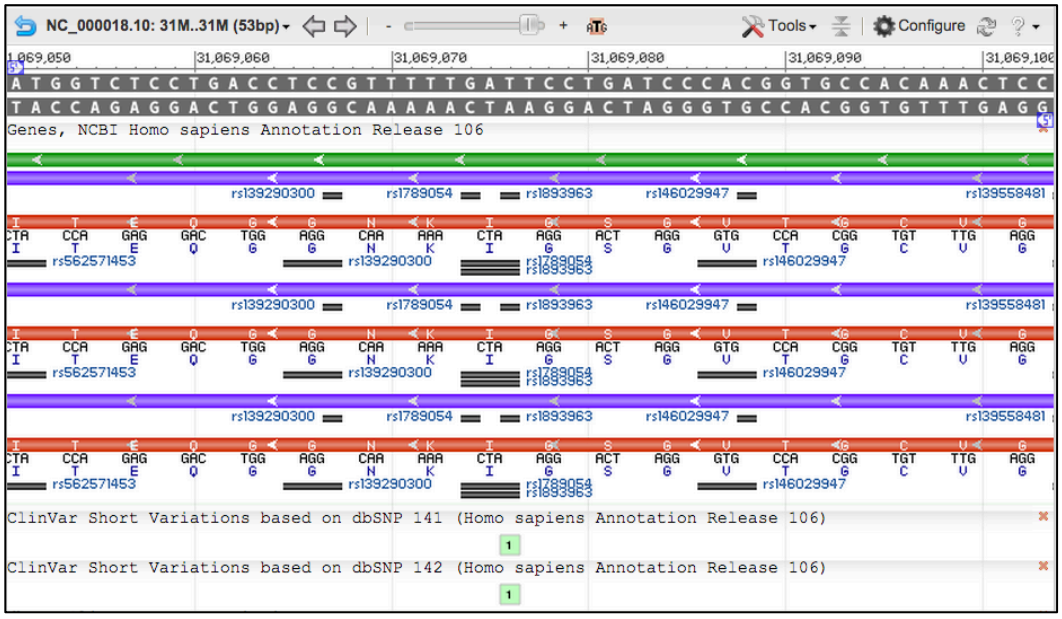
from the 1000Genomes project data, a different set of populations from all over the world. Select minor allele frequency >=0.05.



This leaves only two variants in the table.

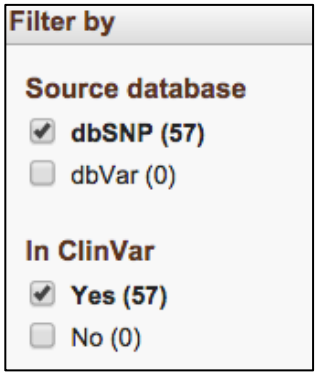


Take a look at the missense variant rs1893963 with a minor allele frequency of 19%. You can link to this variant in the viewer by clicking on the location. Notice that there are overlapping variants in this region, one that affects a different position of in the codon and results in a different amino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

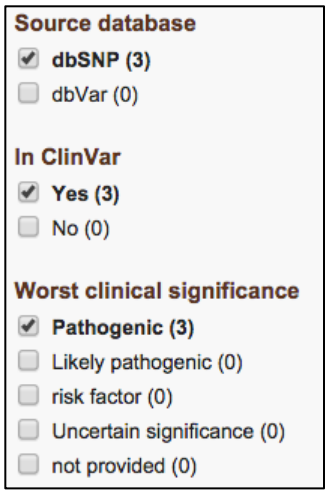


Go back to the Exon/Gene navigator and restore the full-view as before by selecting "Go to gene (with pad)" in the "Region" menu. Now you can find some pathogenic

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



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".



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.

	Variant ID	Location	Variant type	Gene	Molecular consequences	Worst clinical significance	1000G MAF	GO-ESP MAF	Publications
▸	rs397514043	31,074,730	deletion	DSC2	frameshift variant	Pathogenic			
▾	rs397514041	31,080,186	deletion	DSC2	frameshift variant	Pathogenic			

Alleles associated with rs397514041

Allele information					ClinVar information				
Variant allele	Transcript change	RefSeq	Protein change	Molecular consequence	Condition	Worst clinical significance	Submitters	Highest review status	Last evaluated
delG	c.1430delC	NM_004949.3	Thr477Metfs	frameshift variant	Arrhythmogenic right ventricular cardiomyopathy, type 11	Pathogenic	1	classified by single submitter	Sep 03, 2013
delG	c.1430delC	NM_024422.3	Thr477Metfs	frameshift variant	Arrhythmogenic right ventricular cardiomyopathy, type 11	Pathogenic	1	classified by single submitter	Sep 03, 2013
delG	c.1001delC	XM_005258206.2	Thr334Metfs	frameshift variant	Arrhythmogenic right ventricular cardiomyopathy, type 11	Pathogenic	1	classified by single submitter	Sep 03, 2013

▸	rs397514042	31,087,815	single nucleotide variant	DSC2	splice acceptor variant	Pathogenic			

The expanded view shows the HGVS expression and consequences for each of the three splice variants. The variant is asserted to be causative for "Arrhythmogenic right ventricular cardiomyopathy, type 11", a particular syndrome. Follow the "Pathogenic" link to ClinVar.

NM_004949.3(DSC2):c.1430delC (p.Thr477Metfs)

NM_004949.3(DSC2):c.1430delC (p.Thr477Metfs) Go to:

Variant type: Deletion

Cytogenetic location: 18q12.1

Genomic location: Chr18:28660152 (on Assembly GRCh37)
Chr18:31080186 (on Assembly GRCh38)

HGVS: NG_008208.2:g.27240delC
NM_004949.3:c.1430delC
NC_000018.10:g.31080186delG
...more

Links: OMIM: 125645.0001
dbSNP: 397514041

NCBI 1000 Genomes Browser: rs397514041

Molecular consequence: NM_004949.3:c.1430delC: frameshift variant [Sequence Ontology SO:0001589]

Clinical significance

NM_004949.3(DSC2):c.1430delC (p.Thr477Metfs) Help

Clinical significance: Pathogenic/Likely pathogenic
Review status: ★ ☆ ☆ ☆ (1/4)

Number of submission(s): 1

Condition(s)

Arrhythmogenic right ventricular cardiomyopathy, type 11 [MedGen - OMIM]

See supporting ClinVar records

1 Affected Gene

desmocollin 2 (DSC2) [Gene - OMIM - Variation viewer]

Search ClinVar for variants within DSC2

Search ClinVar for variants including DSC2

Browser views

RefSeqGene

Variation viewer [GRCh38 - GRCh37]

UCSC [GRCh38/hg38 - GRCh37/hg19]

Assertion and evidence details Go to:

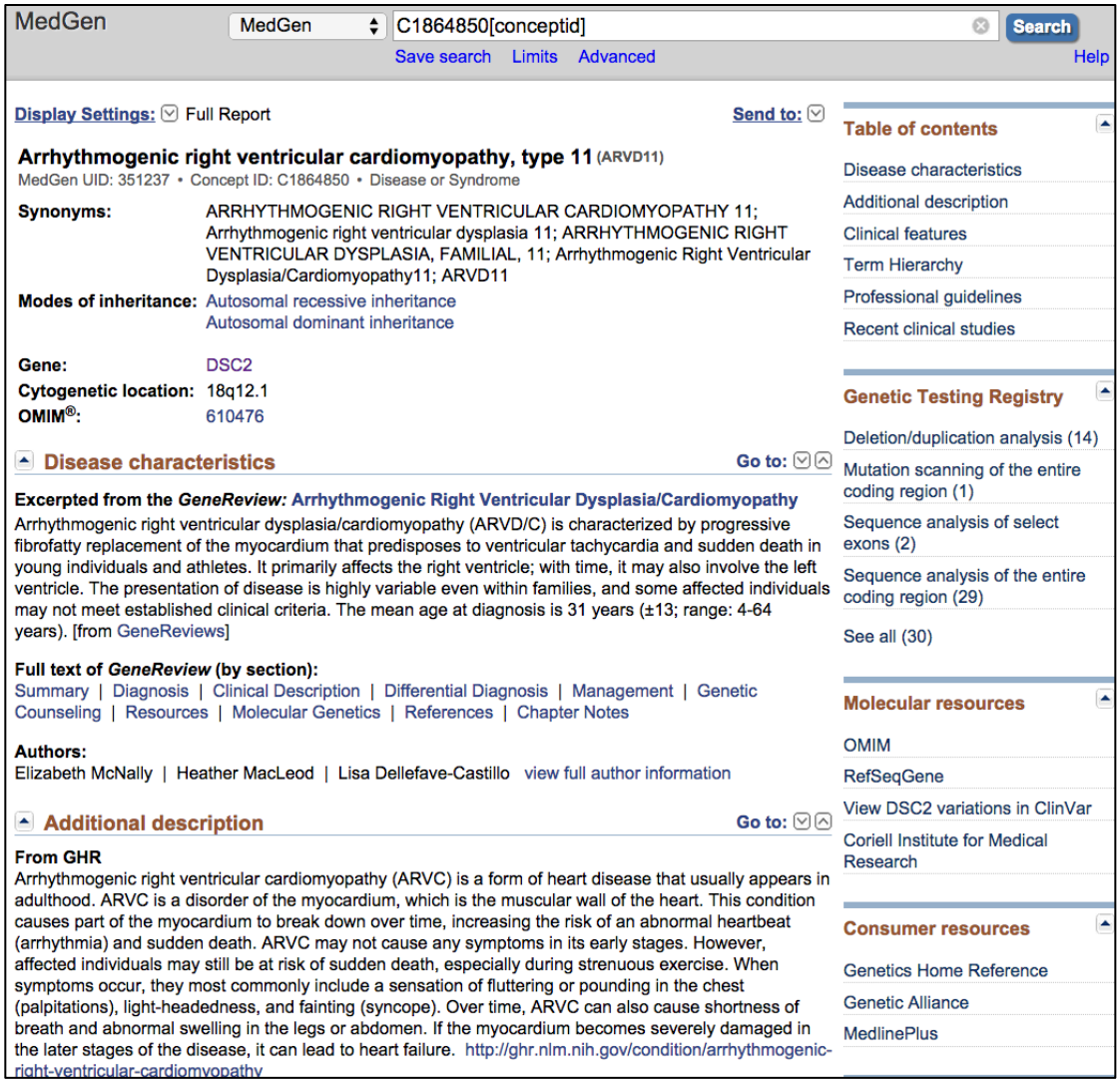
Clinical Assertions | Evidence

Help

Germline

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter (Last submitted)	Submission accession
Pathogenic (Sep 3, 2013)	classified by single submitter (literature only)	literature only	Arrhythmogenic right ventricular cardiomyopathy, type 11 [MedGen | OMIM]	germline	PubMed (2)	OMIM (Dec 30, 2010)	SCV000038621

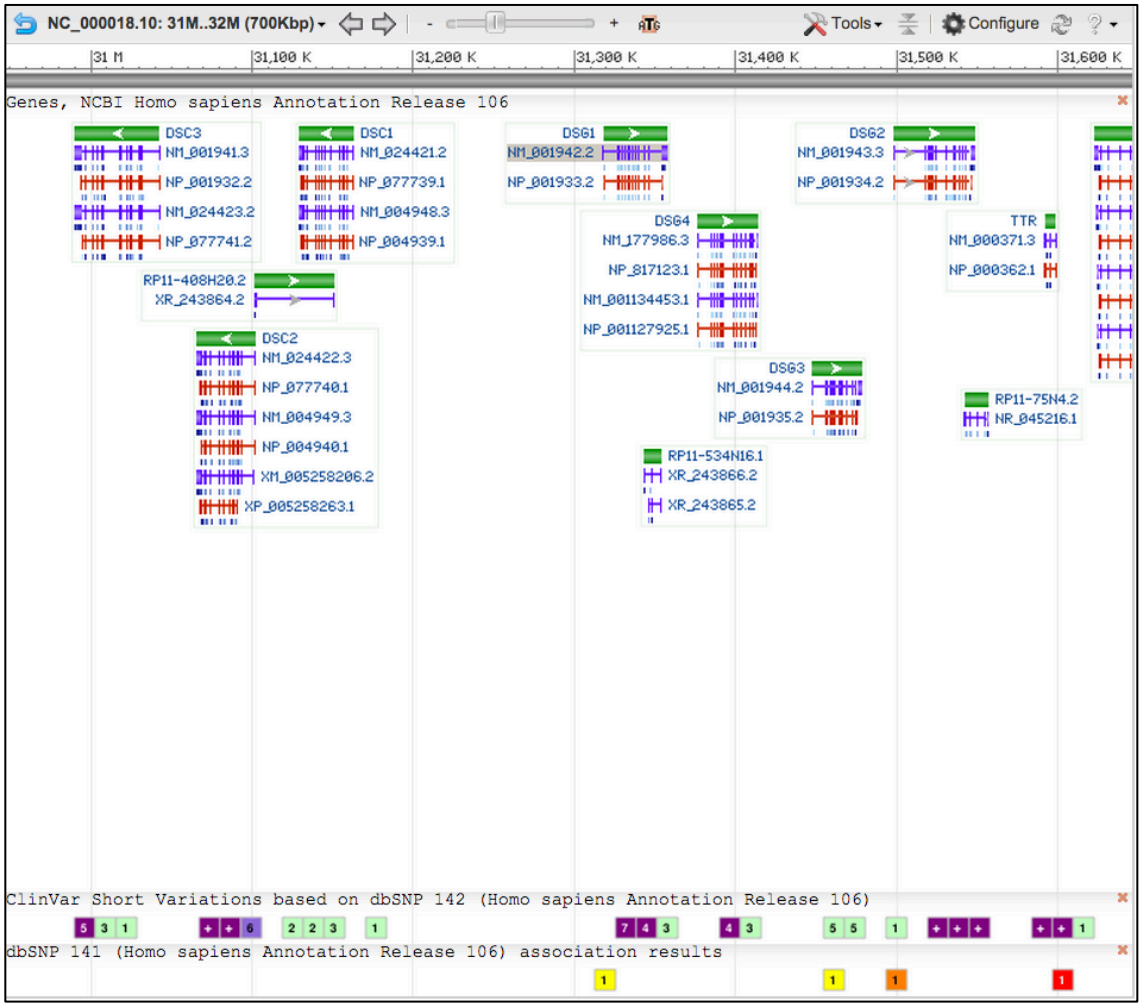
This variant comes from the OMIM record and hence is labeled as literature only. OMIM is only one of the sources of assertions in ClinVar. Follow the link to MedGen on the right to learn more about this particular syndrome. MedGen aggregates data from number of sources and is particularly helpful for 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.



Adding tracks and navigating

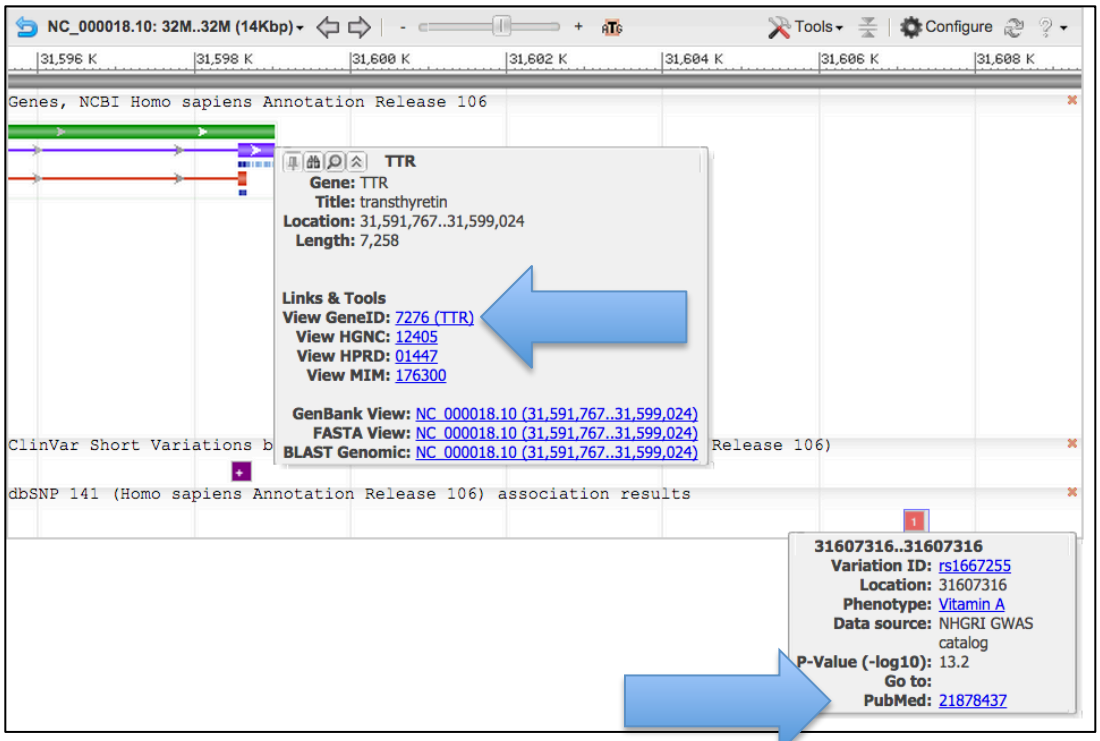
Go back to the Variation Viewer to add an additional track. Click the configure menu in the sequence viewer and select the Variation tracks on the right-hand-side. Check

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).



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.

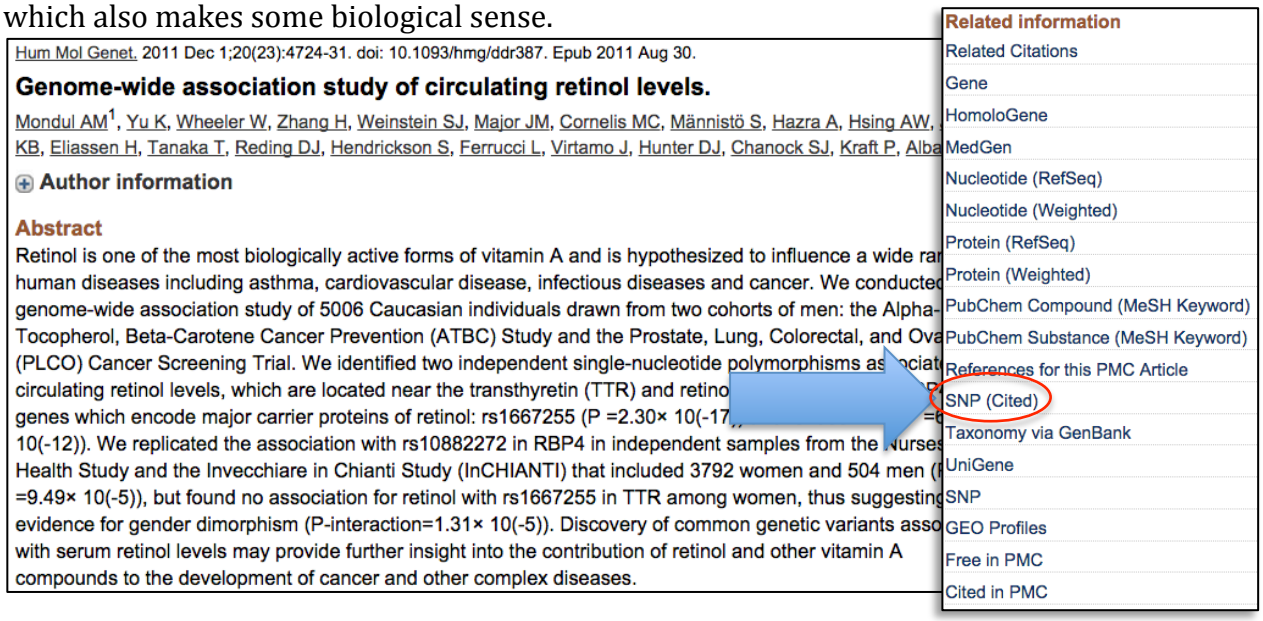
TTR transthyretin [*Homo sapiens* (human)]

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[1], 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, Alba

Author information

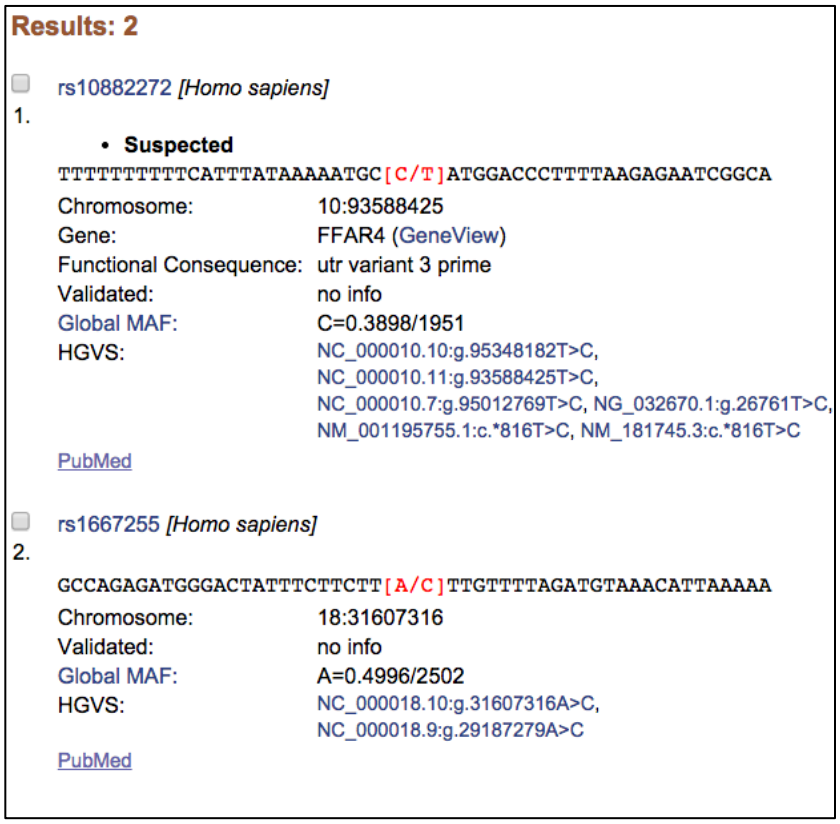
Abstract

Retinol is one of the most biologically active forms of vitamin A and is hypothesized to influence a wide ra human diseases including asthma, cardiovascular disease, infectious diseases and cancer. We conducte 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 Ova (PLCO) Cancer Screening Trial. We identified two independent single-nucleotide polymorphisms as circulating retinol levels, which are located near the transthyretin (TTR) and retino genes which encode major carrier proteins of retinol: rs1667255 (P =2.30× 10(-17 10(-12)). 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 (=9.49× 10(-5)), but found no association for retinol with rs1667255 in TTR among women, thus suggesting evidence for gender dimorphism (P-interaction=1.31× 10(-5)). Discovery of common genetic variants asso 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
- SNP
- 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

1. rs10882272 *[Homo sapiens]*

- **Suspected**

TTTTTTTTTTCATTTATAAAAATGC[C/T]ATGGACCCTTTTAAGAGAATCGGCA

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

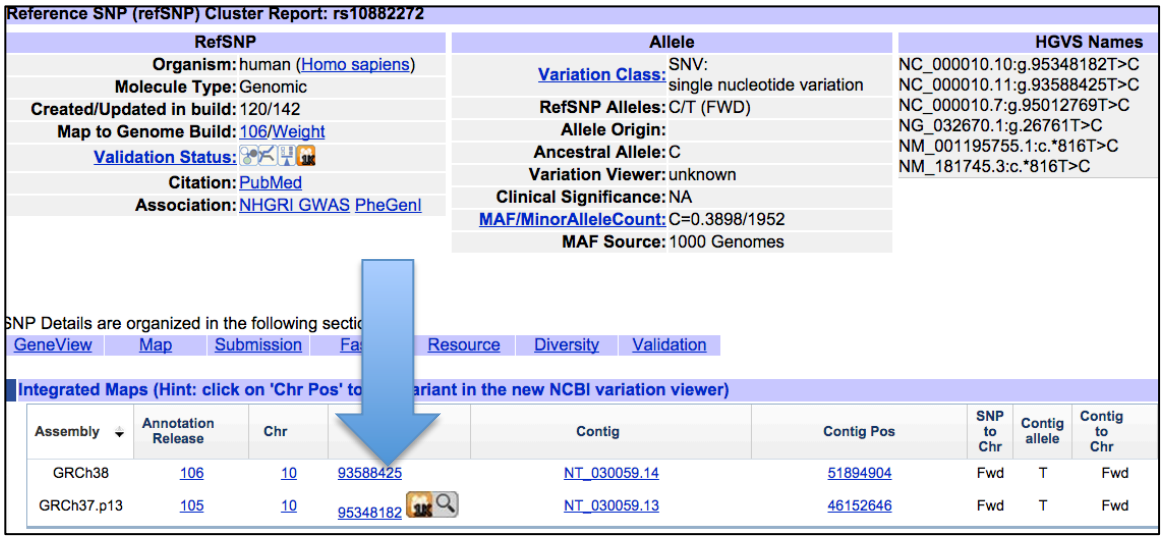
2. rs1667255 *[Homo sapiens]*

GCCAGAGATGGGACTATTTCTTCTT[A/C]TTGTTTTAGATGTAAACATTAAAAA

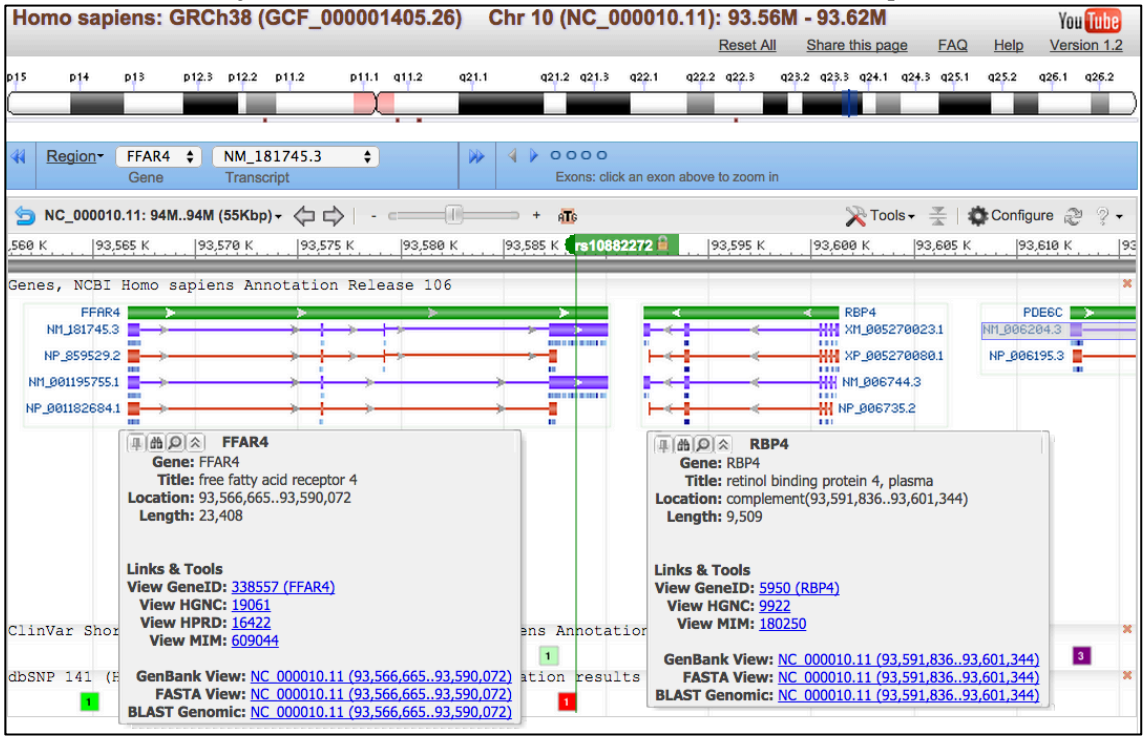
Chromosome:	18:31607316
Validated:	no info
Global MAF:	A=0.4996/2502
HGVS:	NC_000018.10:g.31607316A>C, NC_000018.9:g.29187279A>C

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.



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



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?