

Training materials

- Ensembl training materials are protected by a CC BY license
- <http://creativecommons.org/licenses/by/4.0/>
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers
- <http://www.ensembl.org/info/about/publications.html>



<http://training.ensembl.org/events>



Viewing the data in Ensembl

Benjamin Moore
Ensembl Outreach Officer

Slides available from
training.ensembl.org/events/



Why do we need genome browsers?

1977: 1st genome to be sequenced (5 kb)
2004: finished human sequence (3 Gb)



<http://training.ensembl.org/events>



EMBL-EBI

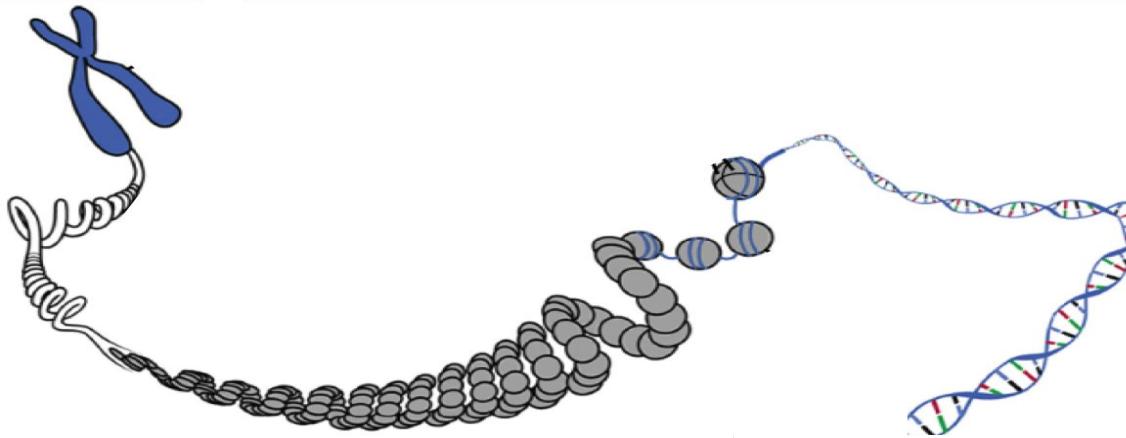


Why do we need genome browsers?

CGGCCTTGGGCTCCGCCTTCAGCTAAGACTTAACCTCCCTCCAGCTGTCCCAGATGACGCCATCTGAAATTCTGGAAACACGATCAC
TTAACGGAATATTGCTTTGGGAAGTGTACAGCTGCTGGCACGCTGTATTGCCTACTTAAGCCCCTGTAATTGCTGTATTGCTGATTG
CGAAGACATGCTGATGGATTACCAAGGGCGGCGTGGCTCTAACTGGAGCCCTGTCCCCACTAGCCACGCGTCACTGGTAGCGTGATTG
GAAACTAAATCGTATGAAAATCCTCTCTAGTCGCACTAGCCACGTTCGAGTGCTTAATGGCTAGTGGCACCGGTTGGACAGCAC
GCTGAAAATGTTCCCATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGACTGAATTAGAACAAATTTCAGCGCTCTGAA
GTTTACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTGGCTTCTGTTGCAGACTTACCAAGCATTGGAGGA
ATATCGTAGGAAAAATGCCTATTGGATCCAAAGAGAGGCCAACATTGGAAATTGAAACAGCCTGCAACAAAGCAGGTATTGACA
ATTTATATAACTTATAAATTACACCGAGAAAGTGTCTAAAAAATGCTGCTAAAACCAGTACGTACAGTGTGCTAGAACCCAT
AAACTGTTCTTATGTGTATAAATCCAGTTAACACATAATCATGTTGCAGGTTAACCATGATAAATATAGAACGTCTAGTGGATA
AAGAGGAAACTGGCCCTTGACTAGCAGTAGGAACAAATTACTAACAAATCAGAACATTATGTTACTTATGGCAGAAGTGTCCA
TTGGTTTCAGTACTCCTTACTCTTAAATGATCTAGGACCCCCGGAGTGCTTTGTTATGTAGCTTACCATATTAGAAATTAAA
AAGAATTAAAGGCTGGCGTGGCTCACGCTGTAATCCCAGCACTTGGAGGCCAGGTGGCGGATCATTGAGGCCAGAAGTTGA
GACCAGCCTGGCCAACATGGTGAACACCTATCTACTAAAATACAAAAATGTGCTGCGTGTGGTGGCGCTGTAATCCAGCTAC
ACGGGAGGTGGAGGCAGGAGAACGCTGAACCCCTGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCAC
TAGCATGACTCTGTCTAAAACAAACAAACAAAAACTAACAGATTAAAGTTAACATTAAAGTAAAGCTAACCATTGCA
TATTATCACAACATTCTAGGAAAAAAACTTTGAAAACAAGTGAATGAGTGGAAATAGTTTACATTGCA
AAATAGAGATAGCTGGATTCACTTATCTGTGTAATCTGTTAGGTTAGAAGTATGTGAAAAAAATTACCTCACGTT
ATATTAAATAGTTTCAGTTACTTTGGTATTTCCTGTACTTGCATAGATTCAAAGATCTAACAGATACCA
CCATGTCGCAACATCATGCAGTGATTATTGAAAGATAGTGGTGTCTGAATTACAAAGTTCAAATATTGATAAATTGCA
TTTAAAAATCTCATTCAATTACACCAGGATGTCAGAAAAGTCTTAAAGATTGGTAGAAATGAGCCACTG
TTGAAAGTTCACATTGTCATTGACAACAAACTGTTCCCTGCAGCAACAGATCACTCATTGAGTTGTGAGAAAATGT
TATTAAAGTTGAAATAACTTGTCAAGTAAAAATGACTTTCAATTGAAAGATAGTGGTGTCTGAATTACAAAGTTCAA
AGTGTGTTCTAGGCAGTATTGACTTCAGTATGCAGAAGTGTGTTATGTATGCTTCAAGTATTGAGGATTATTAAAGAAGT
GCATTGAGCTTCGAAATTAAATTTCATTGCTTCATTAGGACATTCTACATTAAACTGGCATTATTACTATT
TCAGTGGTAAGGAATATAATGGCTACTAGTATTAGTTGGTGCCTGCCACTGCCATAACTCATG
TTGCACTGTTGATAACAAATGTCAACATCATGAAAAAGGTTGAAAAAGGAATATTAAATAGTTCA
<http://training.ensembl.org/events>



Ensembl- unlocking the code

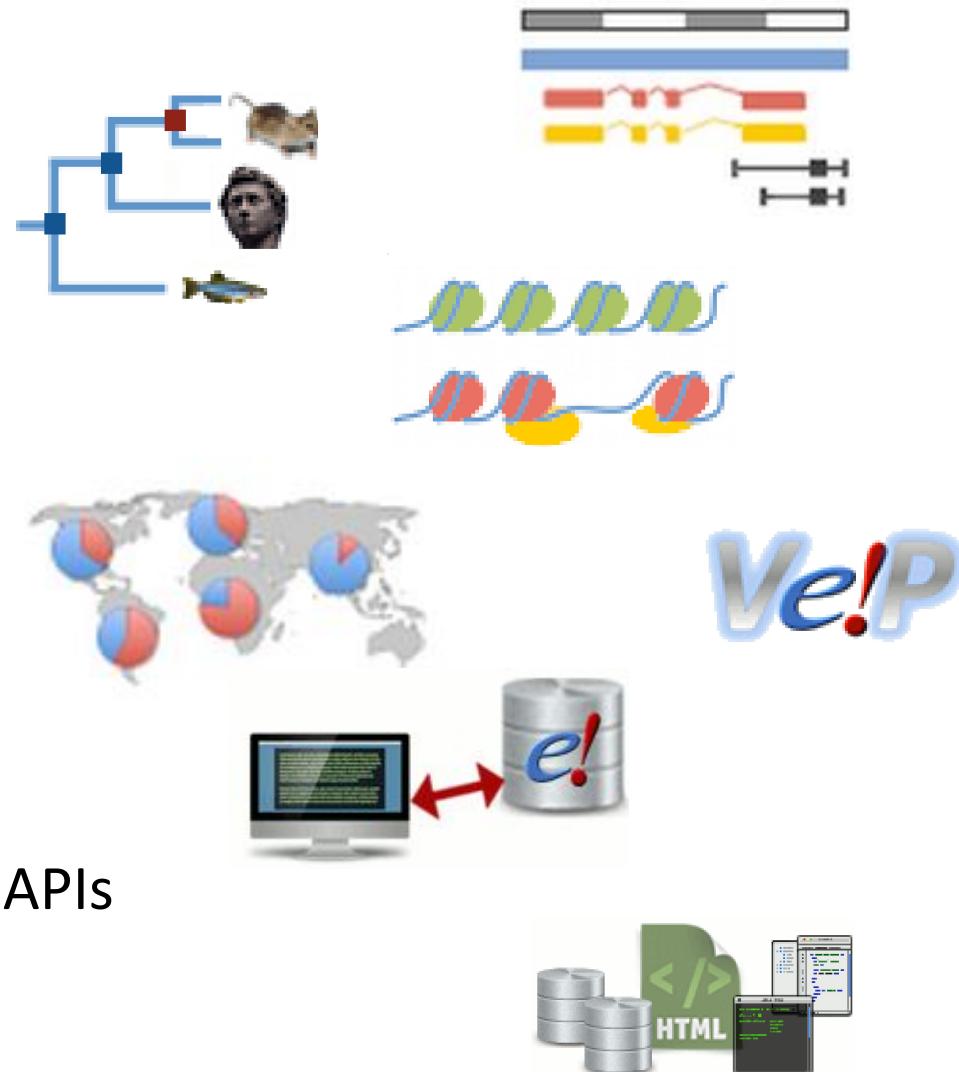


- Genomic assemblies - automated and manual gene annotation
- Variation - Small and large scale sequence variation with phenotype associations
- Comparative Genomics - Whole genome alignments, gene trees
- Regulation - Potential promoters and enhancers, DNA methylation

<http://training.ensembl.org/events>

Ensembl Features

- Gene builds for ~100 species
- Gene trees
- Regulatory build
- Variation display and VEP
- Display of user data
- BioMart (data export)
- Programmatic access via the APIs
- Completely Open Source



<http://training.ensembl.org/events>

Ensembl Genomes- expanding Ensembl



www.ensembl.org

- Vertebrates



- Other representative species

<http://training.ensembl.org/events>



EMBL-EBI



Ensembl Genomes- expanding Ensembl



www.ensembl.org

- Vertebrates



- Other representative species

<http://training.ensembl.org/events>



www.ensemblgenomes.org

- Bacteria



- Fungi



- Protists



- Metazoa



- Plants



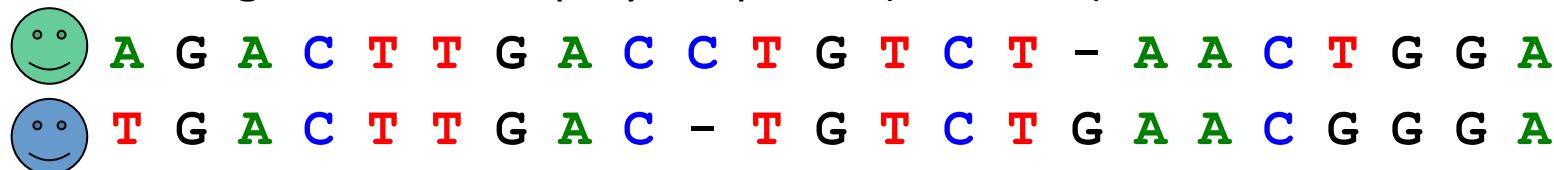
EMBL-EBI



Variation types

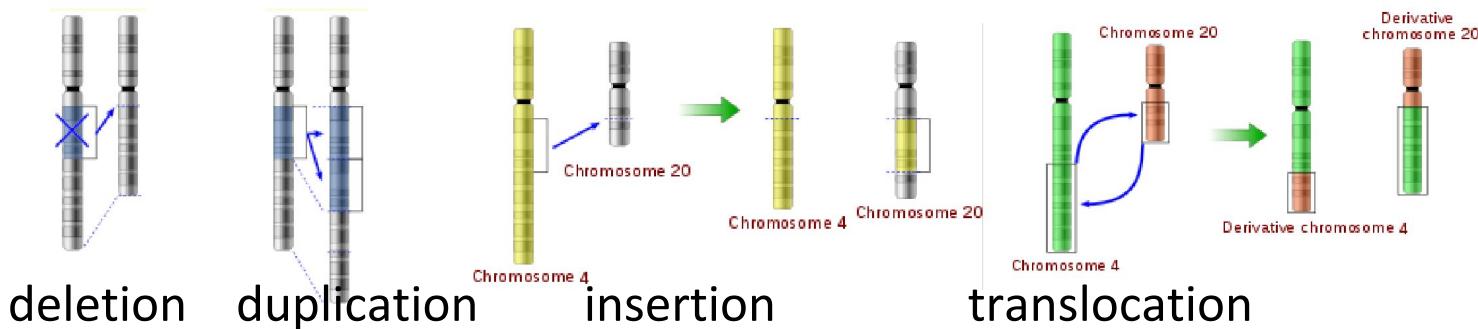
1) Small scale in one or few nucleotides of a gene

- Small insertions and deletions (DIPs or indels)
- Single nucleotide polymorphism (SNP/SNV)



2) Large scale in chromosomal structure (structural variation)

- Copy number variations (CNV)
- Large deletions/duplications, insertions, translocations



23 species with variation data

	Cat <i>Felis catus</i>	3.6 M		Opossum <i>Monodelphis domestica</i>	1.1 M
	Chicken <i>Gallus gallus</i>	24 M		Orangutan <i>Pongo abelii</i>	10 M
	Chimpanzee <i>Pan troglodytes</i>	1.6 M		Pig <i>Sus scrofa</i>	67 M
	Cow <i>Bos taurus</i>	104 M		Platypus <i>Ornithorhynchus anatinus</i>	1.3 M
	Dog <i>Canis familiaris</i>	5.9 M		Rat <i>Rattus norvegicus</i>	5 M
	Fruitfly <i>Drosophila melanogaster</i>	6.7 M		S. cerevisiae <i>Saccharomyces cerevisiae</i>	263 K
	Gibbon <i>Nomascus leucogenys</i>	1.1 M		Sheep <i>Ovis aries</i>	61 M
	Goat <i>Capra hircus</i>	37 M		Tetraodon <i>Tetraodon nigroviridis</i>	902 K
	Horse <i>Equus caballus</i>	21 M		Turkey <i>Meleagris gallopavo</i>	9 K
	Human <i>Homo sapiens</i>	665 M		Zebra Finch <i>Taeniopygia guttata</i>	1.7 M
	Macaque <i>Macaca mulatta</i>	53 M		Zebrafish <i>Danio rerio</i>	17 M
	Mouse <i>Mus musculus</i>	84 M			

<http://training.ensembl.org/events>



EMBL-EBI



Where does the data come from?

The Ensembl variation process



<http://www.ensembl.org/info/genome/variation/index.html>

<http://training.ensembl.org/events>



Ensembl variation process: Import



Import variant data from
publicly available archives
and data repositories

dbSNP
Short Genetic Variations

DGVA^{rchive}



EVA



COSMIC

The core of COSMIC, an expert-curated database of somatic mutations



Cancer Gene Census

A catalogue of genes with mutations that are causally implicated in cancer

http://www.ensembl.org/info/genome/variation/species/sources_documentation.html

<http://training.ensembl.org/events>



EMBL-EBI



Finding your variant

The screenshot shows the Ensembl homepage with a search bar at the top containing 'rs1800566'. A red arrow points from this search result down to the 'What's New in Ensembl Release 90' section.

Search: All species for Go
e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Browse a Genome
Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes
Human GRCh38.p10 Mouse GRCm38.p5

Find a Data Display
TABLE
HEATMAP
SEQUENCE
PIE CHART
Not sure how to find the data visualisation you need? With our new [Find a Data Display](#) page, you can choose a gene, region or variant and then browse a selection of relevant visualisations

What's New in Ensembl Release 90

- New rodent species
- New genome annotation on the pig assembly *Sscrofa11.1*
- Mouse: update to Ensembl-Havana GENCODE gene set
- Update to Ensembl-Havana human GENCODE gene set (release 27)
- New probe mapping data for five new rodents

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

You can find variant information by:

- searching for a gene or phenotype and examining the variant or associated features table
- viewing a genomic location and clicking on a variant in one of the variant tracks
- searching directly by variant name

<http://training.ensembl.org/events>



The Variant Tab

Location: 16:69,710,742-69,711,742 Variant: rs1800566 Jobs ▾

Variant displays

- Explore this variant
- Genomic context
 - Genes and regulation
 - Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

rs1800566 SNP

Most severe consequence
Alleles
Location
Co-located variant
Evidence status ⓘ
HGVS names
Synonyms

Genotyping chips
Original source
About this variant
Description from SNPedia

Explore this variant ⓘ

- Genomic context
- Genes and regulation
- Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

missense variant | See all predicted consequences
G/A | Ancestral: G | MAF: 0.29 (A) | Highest population MAF: 0.50
Chromosome 16:69711242 (forward strand) | VCF: 16 69711242 rs1800566 G A
HGMD-PUBLIC CM950861

This variant has 21 HGVS names - Show +
This variant has 10 synonyms - Hide ⚡

- Archive dbSNP rs4134727 ↗, rs4149351 ↗, rs57135274 ↗
- ClinVar RCV000018301 ↗, RCV000018300 ↗, RCV000211294 ↗, RCV000434090 ↗, RCV000018302 ↗
- LSDB 1560
- Uniprot VAR_008384 ↗

This variant has assays on 12 chips - Show +
Variants (including SNPs and indels) imported from dbSNP (release 150) | View in dbSNP ↗
This variant overlaps 7 transcripts, 1 regulatory feature, has 4674 sample genotypes, is associated with 6 phenotypes and is mentioned in 159 citations.
rs1800566 (C609T, Pro187Ser) is a SNP within NQO1 (NAD(P)H dehydrogenase (quinone 1)). A * at this location denotes the NQO1*2 allele.... Show +

Icons available (grey available)

Icon	Value
association	association
benign	benign
confers sensitivity	confers sensitivity
drug response	drug response
likely benign	likely benign
likely pathogenic	likely pathogenic
not provided	not provided
other	other
pathogenic	pathogenic
protective	protective
risk factor	risk factor
uncertain significance	uncertain significance

Menu of pages,
available on all
pages

Ensembl variation process: QC



- Mapping to reference assembly
 - GRCh37 → GRCh38
- Checks on alleles
- Checks for IUPAC ambiguity codes
- Excluding ‘suspect’ variants

[http://www.ensembl.org/info/genome/variation/prediction/
variant_quality.html#quality_control](http://www.ensembl.org/info/genome/variation/prediction/variant_quality.html#quality_control)

<http://training.ensembl.org/events>



EMBL-EBI



Quality Control

Icon	Name	Description	QC Type	Reported failure reason
	Multiple observations	The variant has multiple independent dbSNP submissions, i.e. submissions with a different source and different discovery samples	Mapping checks	Variant does not map to the genome
	Frequency	The variant is reported to be polymorphic in at least one sample		Variant maps to more than 1 location
	Cited	The variant is cited in a PubMed article.		
	Phenotype or Disease	The variant is associated with at least one phenotype or disease.		
	1000 Genomes	The variant was discovered in the 1000 Genomes Project (human only)		
	ExAC	The variant was discovered in the Exome Aggregation Consortium (human only).		
	HapMap	The variant is polymorphic in at least one HapMap panel (human only)		Mapped position is not compatible with reported alleles
	ESP	The variant was discovered in the Exome Sequencing Project (human only).		None of the variant alleles match the reference allele
<p>We summarise the information supporting each dbSNP variant to give a guide to its reliability</p> <p>We run basic checks and flag suspicious variants</p>				
<p>rs14390 SNP</p> <p> This variation has been flagged</p> <ul style="list-style-type: none"> Flagged as suspect by dbSNP Variation maps to 2 genomic locations <p>Select a location: <input type="button" value="None selected"/> <input type="button" value="Go"/></p> <p>Original source Variants (including SNPs and indels) imported from dbSNP (release 138) Alleles C/T Ancestral: T Ambiguity code: Y Location This variation maps to 2 genomic locations; None selected Evidence status </p>				
<p></p> <p></p>				
<p>Checks on the alleles of refSNPs</p> <p>Loci with no observed variant alleles in dbSNP</p> <p>Alleles contain ambiguity codes</p> <p>Alleles contain non-nucleotide characters</p>				
<p>Checks on the alleles in dbSNP submissions</p> <p>Additional submitted allele data from dbSNP does not agree with the dbSNP refSNP alleles</p>				
<p>External failure classification</p> <p>Flagged as suspect by dbSNP</p>				

<http://training.ensembl.org/events>



Ensembl variation process: Linked data



Import 'accessory' data

- Allele frequencies
- Phenotype/disease
- Publication data



ClinVar



Allele Frequencies- the 1000 Genomes Project

Name	Size	Description
1000GENOMES:phase_3:ALL	2504	All phase 3 individuals
1000GENOMES:phase_3:AFR	661	African
• 1000GENOMES:phase_3:ACB	96	African Caribbean in Barbados
• 1000GENOMES:phase_3:ASW	61	African Ancestry in Southwest US
• 1000GENOMES:phase_3:ESN	99	Esan in Nigeria
• 1000GENOMES:phase_3:GWD	113	Gambian in Western Division, The Gambia
• 1000GENOMES:phase_3:LWK	99	Luhya in Webuye, Kenya
• 1000GENOMES:phase_3:MSL	85	Mende in Sierra Leone
• 1000GENOMES:phase_3:YRI	108	Yoruba in Ibadan, Nigeria
1000GENOMES:phase_3:AMR	347	American
• 1000GENOMES:phase_3:CLM	94	Colombian in Medellin, Colombia
• 1000GENOMES:phase_3:MXL	64	Mexican Ancestry in Los Angeles, California
• 1000GENOMES:phase_3:PEL	85	Peruvian in Lima, Peru
• 1000GENOMES:phase_3:PUR	104	Puerto Rican in Puerto Rico
1000GENOMES:phase_3:EAS	504	East Asian
• 1000GENOMES:phase_3:CDX	93	Chinese Dai in Xishuangbanna, China
• 1000GENOMES:phase_3:CHB	103	Han Chinese in Beijing, China
• 1000GENOMES:phase_3:CHS	105	Southern Han Chinese, China
• 1000GENOMES:phase_3:JPT	104	Japanese in Tokyo, Japan
• 1000GENOMES:phase_3:KHV	99	Kinh in Ho Chi Minh City, Vietnam

1000GENOMES:phase_3:EUR	503	European
• 1000GENOMES:phase_3:CEU	99	Utah residents with Northern and Western European ancestry
• 1000GENOMES:phase_3:FIN	99	Finnish in Finland
• 1000GENOMES:phase_3:GBR	91	British in England and Scotland
• 1000GENOMES:phase_3:IBS	107	Iberian populations in Spain
• 1000GENOMES:phase_3:TSI	107	Toscani in Italy
1000GENOMES:phase_3:SAS	489	South Asian
• 1000GENOMES:phase_3:BEB	86	Bengali in Bangladesh
• 1000GENOMES:phase_3:GIH	103	Gujarati Indian in Houston, TX
• 1000GENOMES:phase_3:ITU	102	Indian Telugu in the UK
• 1000GENOMES:phase_3:P JL	96	Punjabi in Lahore, Pakistan
• 1000GENOMES:phase_3:STU	102	Sri Lankan Tamil in the UK

The 1000 Genomes Phase 3 data provides genotype data for 2504 individuals from 26 different narrowly defined populations in 5 groupings.

<http://training.ensembl.org/events>

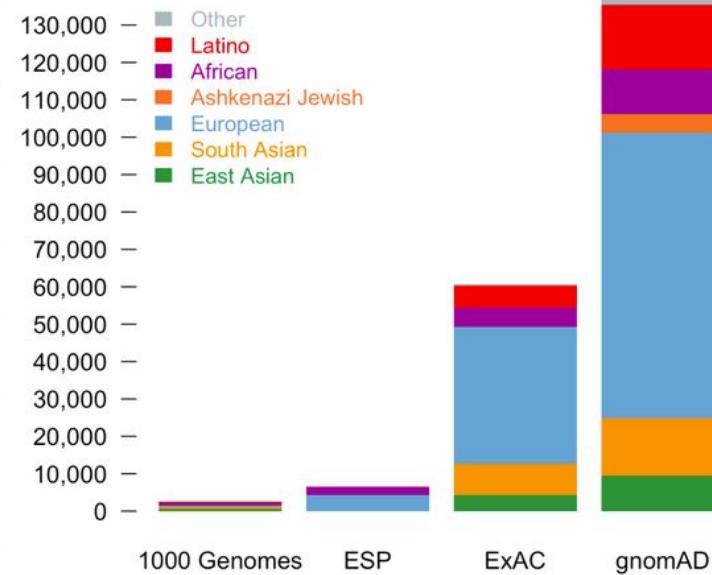


Allele Frequencies- GnomAD

The Genome Aggregation Database provides allele frequency data for over 130,000 samples from 7 different populations

POPULATION	DESCRIPTION	GENOMES	EXOMES	TOTAL
AFR	African/African American	4,368	7,652	12,020
AMR	Admixed American	419	16,791	17,210
ASJ	Ashkenazi Jewish	151	4,925	5,076
EAS	East Asian	811	8,624	9,435
FIN	Finnish	1,747	11,150	12,897
NFE	Non-Finnish European	7,509	55,860	63,369
SAS	South Asian	0	15,391	15,391
OTH	Other (population not assigned)	491	2,743	3,234
Total		15,496	123,136	138,632

Sample numbers



From: <https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/>

<http://training.ensembl.org/events>



Allele Frequencies

Location: 16:69,710,742-69,711,742 Variant: rs1800566 Jobs ▾

Variant displays

- Explore this variant
- Genomic context
 - Genes and regulation
 - Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

rs1800566 SNP

Most severe consequence: missense variant | See all predicted consequences

G/A | Ancestral: G | MAF: 0.29 (A) | Highest population MAF: 0.50

Chromosome 16:69,712,42 (forward strand) | VCF: 16:69,712,42 rs1800566 G A

HGMD-PUBLIC CM950861

This variant has 21 HGVS names - [Show](#) +

This variant has 10 synonyms - [Hide](#) ▾

- Archive dbSNP [rs4134727](#), [rs4149351](#), [rs57135274](#)
- ClinVar [RCV000018301](#), [RCV000018300](#), [RCV000211294](#), [RCV000434090](#), [RCV000018302](#)
- LSDB 1560
- Uniprot [VAR_008384](#)

This variant has assays on 12 chips - [Show](#) +

Variants (including SNPs and indels) imported from dbSNP (release 150) | [View in dbSNP](#)

This variant overlaps 7 transcripts, 1 regulatory feature, has 4674 sample genotypes, is associated with 6 phenotypes and is mentioned in 159 citations.

rs1800566 (C609T, Pro187Ser) is a SNP within NQO1 (NAD(P)H dehydrogenase (quinone 1)). A Ser (T) at this location denotes the NQO1*2 allele.... [Show](#) +

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Genotyping chips

Original source

About this variant

Description from SNPedia

Explore this variant

- Genomic context
- Genes and regulation
- Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

A red arrow points from the "Population genetics" card to the text "Concise summary".

Concise summary

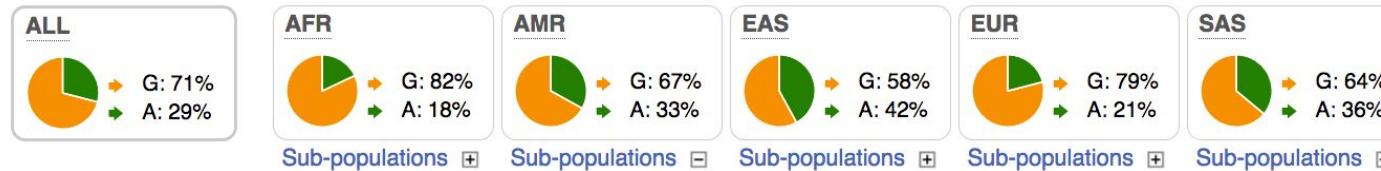
<http://training.ensembl.org/events>



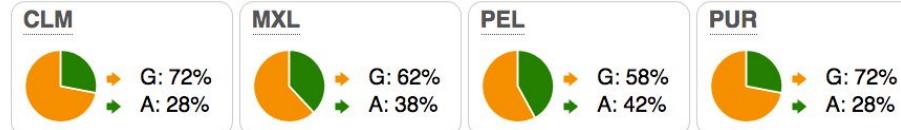
Allele Frequency Distributions

Population genetics ?

1000 Genomes Project Phase 3 allele frequencies



AMR sub-populations



gnomAD exomes (9)

The figure is a screenshot of a table showing gnomAD exomes (9) allele frequency data for various populations. The table includes columns for Population and Allele: frequency (count).

Population	Allele: frequency (count)
gnomADe:ALL	G: 0.748 (181971) A: 0.252 (61327)
gnomADe:AFR	G: 0.811 (12400) A: 0.189 (2892)
gnomADe:AMR	G: 0.608 (20373) A: 0.392 (13153)
gnomADe:ASJ	G: 0.815 (8013) A: 0.185 (1817)
gnomADe:EAS	G: 0.543 (9367) A: 0.457 (7875)
gnomADe:FIN	G: 0.817 (18192) A: 0.183 (4070)
gnomADe:NFE	G: 0.811 (88426) A: 0.189 (20600)
gnomADe:OTH	G: 0.766 (4178) A: 0.234 (1278)
gnomADe:SAS	G: 0.686 (21022) A: 0.314 (9642)

Pie charts and tables display frequency data for different studies. These include TOPMed and UK10K (ALSPAC and TWINS UK) for human, WTSI Mouse Genomes Project and NextGen

<http://training.ensembl.org/events>



EMBL-EBI



Phenotype and Disease Data

ClinVar

European
eGAP
archive



OMIM®

DECIPHER
GRCh37

DGVA archive



IMPC
INTERNATIONAL MOUSE
PHENOTYPING CONSORTIUM



orphanet

LOVD
Leiden Open Variation Database

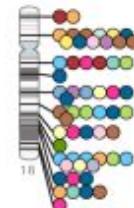
dbGaP
GENOTYPES and PHENOTYPES



ZFIN

OMIA - ONLINE MENDELIAN INHERITANCE IN ANIMALS

NHGRI-EBI
GWAS
Catalog



Variant

Gene

Structural Variant

QTL

Attributes types held for phenotype features include:

- clinical significance reported by ClinVar
- inheritance type
- reported genes /variants
- risk allele
- p-value
- odds ratio
- beta coefficient

<http://training.ensembl.org/events>

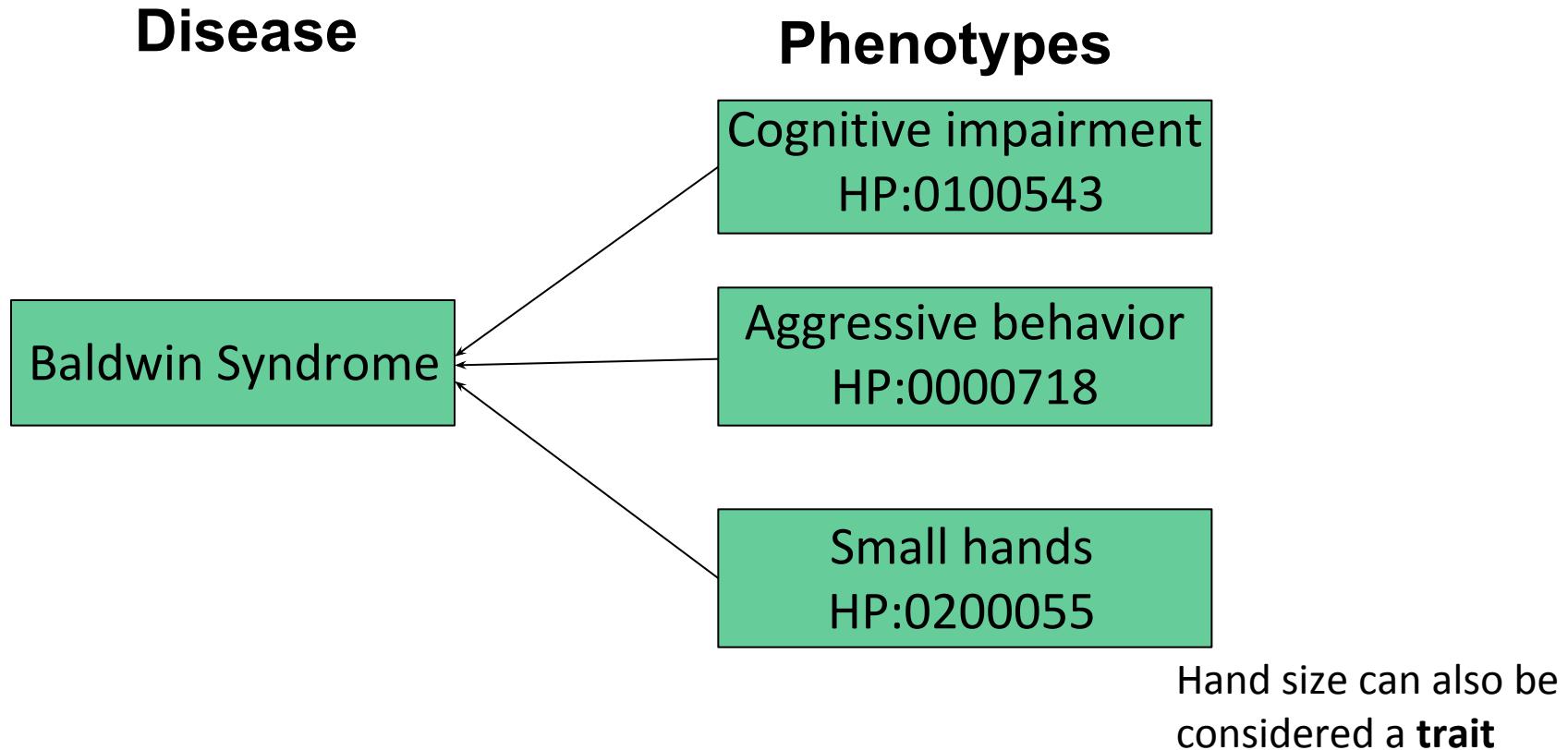


EMBL-EBI



Terminology

A disease can be considered as a bundle of phenotypes observed in the subjects, often named after the person who studied and characterised it.



Viewing Ontology Mappings by Variant

Additional columns in the table of diseases linked to a variant show the Ontology terms we have mapped to the description

It is now clear these descriptions refer to the same disease

Show/hide columns (2 hidden)	Filter					
Phenotype, disease and trait	Source(s)	Mapped Terms	Ontology Accessions	Study	Clinical significance	Reported gene(s)
DARIER DISEASE	Uniprot	Darier disease	Orphanet:218	MIM:124200	-	ATP2A2
DARIER DISEASE	OMIM	Darier disease	Orphanet:218	MIM:108740	-	ATP2A2
Keratosis follicularis	ClinVar	Darier disease, Darier's disease	EFO:0004124, Orphanet:218	-		ATP2A2

Phenotype and Disease Data

Location: 16:69,710,742-69,711,742 Variant: rs1800566 Jobs ▾

Variant displays

- Explore this variant**
- Genomic context
 - Genes and regulation
 - Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

rs1800566 SNP

Most severe consequence missense variant | See all predicted consequences

Alleles G/A | Ancestral: G | MAF: 0.29 (A) | Highest population MAF: 0.50

Location Chromosome 16:69711242 (forward strand) | VCF: 16 69711242 rs1800566 G A

Co-located variant HGMD-PUBLIC CM950861

Evidence status ⓘ Clinical significance ⓘ

HGVS names

Synonyms

Genotyping chips

Original source

About this variant

Description from SNPedia

This variant has 21 HGVS names - [Show](#) +

This variant has 10 synonyms - [Hide](#) □

- Archive dbSNP [rs4134727](#), [rs4149351](#), [rs57135274](#)
- ClinVar [RCV00018301](#), [RCV00018300](#), [RCV000211294](#), [RCV000434090](#), [RCV00018302](#)
- LSDB 1560
- Uniprot [VAR_008384](#)

This variant has assays on 12 chips - [Show](#) +

Variants (including SNPs and indels) imported from dbSNP (release 150) | [View in dbSNP](#)

This variant overlaps 7 transcripts, 1 regulatory feature, has 4674 sample genotypes, is associated with 6 phenotypes and is mentioned in 159 citations.

rs1800566 (C609T, Pro187Ser) is a SNP within NQO1 (NAD(P)H dehydrogenase (quinone 1)). A Ser (T) at this location denotes the NQO1*2 allele.... [Show](#) +

Explore this variant ⓘ

- Genomic context
- Genes and regulation
- Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

Very concise summary

<http://training.ensembl.org/events>



Phenotype and Disease Data

Phenotype Data

Significant association(s)

Phenotype, disease and trait								Show/hide columns	Filter
Phenotype, disease and trait	Source(s)	Mapped Terms	Ontology Accessions	Study	Clinical significance	Reported gene(s)	Associated allele		
BENZENE TOXICITY, SUSCEPTIBILITY TO	OMIM	-	-	MIM:125860	-	NQO1	0001		
Alkylating Agents, anthracyclines and related substances, fluorouracil, and Platinum compounds response - Efficacy	ClinVar	-	-	-	    	NQO1	A		
Lung Cancer	ClinVar	lung adenocarcinoma, lung carcinoma, squamous cell carcinoma	EFO:0000571 , EFO:0000707 , EFO:0001071	-	    	NQO1	A		
BENZENE TOXICITY, SUSCEPTIBILITY TO	ClinVar	-	-	-	    	NQO1	A		
Leukemia, post-chemotherapy, susceptibility to	ClinVar	leukemia	EFO:0000565	-	    	NQO1	A		
Breast cancer, post-chemotherapy poor survival in	ClinVar	breast carcinoma	EFO:0000305	-	    	NQO1	A		

Citation Data - Sources

dbSNP

- Publication information is submitted with the variant submission.

EuropePMC

- Publications for which the full text is freely available in PubMed Central are mined for refSNP identifiers.

UCSC Genocoding Project

- Publications for which the text is freely available or those from publishers who have agreed to data mining are mined for refSNP identifiers.

Species	Citations
sheep	31
horse	58
pig	73
rat	90
dog	200
chicken	414
cow	1,625
mouse	8,309
human	594,364

Citation Data

Location: 16:69,710,742-69,711,742 Variant: rs1800566 Jobs ▾

Variant displays

- Explore this variant
- Genomic context
 - Genes and regulation
 - Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

rs1800566 SNP

Most severe consequence missense variant | See all predicted consequences

Alleles G/A | Ancestral: G | MAF: 0.29 (A) | Highest population MAF: 0.50

Location Chromosome 16:69711242 (forward strand) | VCF: 16 69711242 rs1800566 G A

Co-located variant HGMP TRI HUH19 CM950861

Evidence status ⓘ Clinical significance ⓘ

HGVS names

Synonyms

Genotyping chips

Original source

About this variant

Description from SNPedia

This variant has 21 HGVS names - [Show](#) +

This variant has 10 synonyms - [Hide](#) □

- Archive dbSNP [rs4134727](#), [rs4149351](#), [rs57135274](#)
- ClinVar [RCV00018301](#), [RCV00018300](#), [RCV000211294](#), [RCV000434090](#), [RCV00018302](#)
- LSDB 1560
- Uniprot [VAR_008384](#)

This variant has assays on 12 chips - [Show](#) +

Variants (including SNPs and indels) imported from dbSNP (release 150) | [View in dbSNP](#)

This variant overlaps 7 transcripts, 1 regulatory feature, has 4674 sample genotypes, is associated with 6 phenotypes and is mentioned in 159 citations.

rs1800566 (C609T, Pro187Ser) is a SNP within NQO1 (NAD(P)H dehydrogenase (quinone 1)). A Ser (T) at this location denotes the NQO1*2 allele.... [Show](#) +

Explore this variant ⓘ

- Genomic context
- Genes and regulation
- Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes 4674
- Linkage disequilibrium
- Phylogenetic context
- Citations 159

Very concise summary

<http://training.ensembl.org/events>



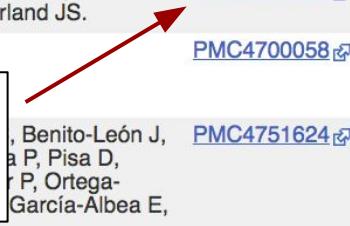
Citation Data

Citations

rs1800566 is mentioned in the following publications

Year	PMID	Title	Author(s)	Full text
2017	28291250	Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism.	Medina-Carmona E, Neira JL, Salido E, Fuchs JE, Palomino-Morales R, Timson DJ, Pey AL.	PMC5349528
2017	28529598	Biological interaction of cigarette smoking on the association between genetic polymorphisms involved in inflammation and the risk of lung cancer: A case-control study in Japan.	Yamamoto Y, Kiyohara C, Suetsugu-Ogata S, Hamada N, Nakanishi Y.	PMC5431513
2017	28282928	Oxidative Stress: A New Target for Pancreatic Cancer Prognosis and Treatment.	Martinez-Useros J, Li W, Cabeza-Morales M, Garcia-Foncillas J.	PMC5372998
2016	26801900	Pharmacogenetics driving personalized medicine: analysis of genetic polymorphisms related to breast cancer medications in Italian isolated populations.	Cocca M, Bedognetti D, La Bianca M, Gasparini P, Girotto G.	PMC4722680
2016	27015811	Associations of air pollution exposure with blood pressure and heart rate variability are modified by oxidative stress genes: A repeated-measures panel among elderly urban residents.	Kim KN, Kim JH, Jung K, Hong YC.	PMC4807581
2016	27019599	Association between L55M polymorphism in Paraoxonase 1 and cancer risk: a meta-analysis based on 21 studies.	Chen L, Lu W, Fang L, Xiong H, Wu X, Zhang M, Wu S, Yu D.	PMC4786067
2016	26873362	Gene Polymorphism Association with Type 2 Diabetes and Related Gene-Gene and Gene-Environment Interactions in a Uyghur Population.	Xiao S, Zeng X, Fan Y, Su Y, Ma Q, Zhu J, Yao H.	PMC4755665
2016	26426434	Necrotizing Enterocolitis Is Not Associated With Sequence Variants in Antioxidant Response Genes in Premature Infants.	Sampath V, Helbling D, Menden H, Dimmock D, Mulrooney NP, Murray JC, Dagle JM, Garland JS.	PMC5055643
2016	26445852	Genetic Biomarkers of Barrett's Esophagus Susceptibility and Progression to Dysplasia and Cancer: A Systematic Review and Meta-Analysis.	Findlay JM, Middleton MR, Tomlinson I.	PMC4700058
2016	26868429	Heme Oxygenase-1 and 2 Common Genetic Variants and Risk for Multiple Sclerosis.	Agúndez JA, García Millán-Pascual J, Díaz Turpín-Fenoll L, Alcolea Cubero S, Ayuso-Fernández A, Benito-León J, Martínez P, Pisa D, Martínez P, Ortega-García J, García-Albea E, Plaza-Nieto JF, Jiménez-Jiménez FJ.	PMC4751624

Link to
full text

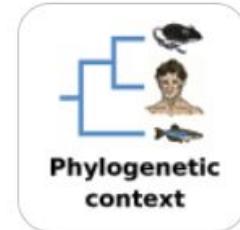


Ensembl variation process: Analysis



Ensembl predicts:

- Variant consequences
- Protein function prediction
- Linkage disequilibrium data
- Variant conservation across species



<http://www.ensembl.org/info/genome/variation/prediction/index.html>

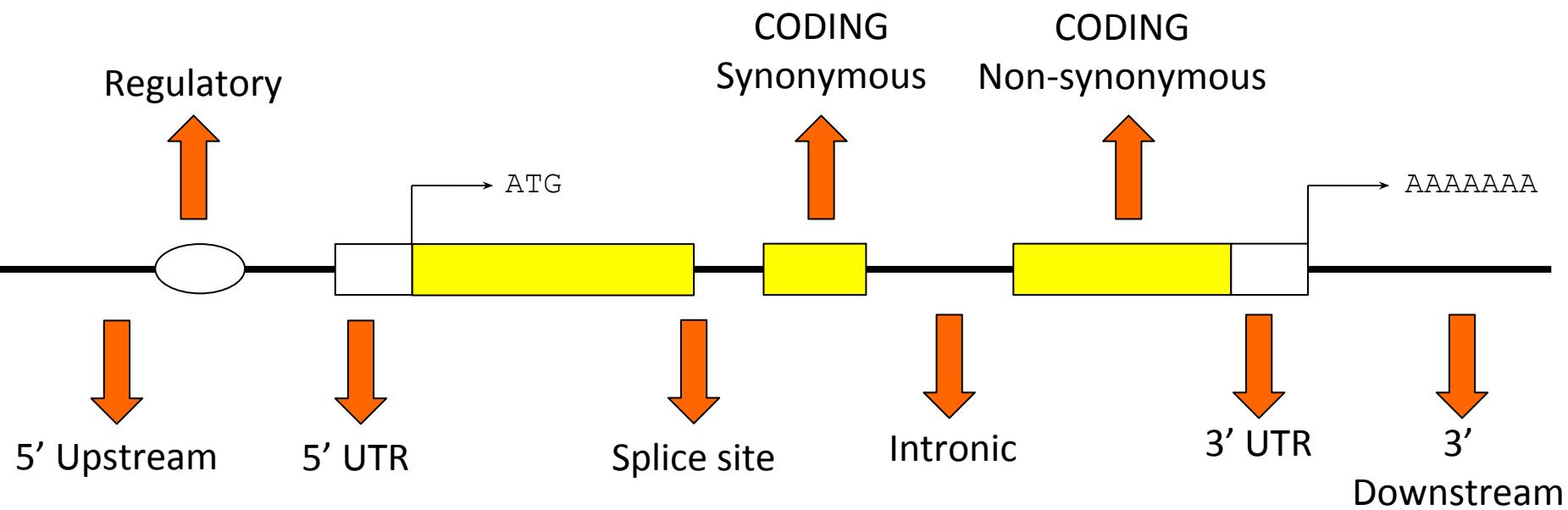
<http://training.ensembl.org/events>



EMBL-EBI



Variation consequences



Consequence terms

* SO term	SO description	SO accession	Old Ensembl term
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Essential splice site
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift coding
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost
initiator_codon_variant	A codon variant that changes at least one base of the first codon of a transcript	SO:0001582	Non synonymous coding
inframe_insertion	An inframe non synonymous variant that inserts bases into the coding sequence	SO:0001821	
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice site
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO:0001626	Partial codon
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid	SO:0001819	Synonymous coding
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains	SO:0001567	
coding_sequence_variant	A sequence variant that changes the coding sequence	SO:0001580	Coding unknown
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA	SO:0001620	Within mature miRNA
5_prime_UTR_variant	A UTR variant of the 5' UTR	SO:0001623	5prime UTR
3_prime_UTR_variant	A UTR variant of the 3' UTR	SO:0001624	3prime UTR
intron_variant	A transcript variant occurring within an intron	SO:0001627	Intronic
NMD_transcript_variant	A variant in a transcript that is the target of NMD	SO:0001621	NMD transcript
non_coding_exon_variant	A sequence variant that changes non-coding exon sequence	SO:0001792	Within non coding gene
nc_transcript_variant	A transcript variant of a non coding RNA	SO:0001619	
upstream_gene_variant	A sequence variant located 5' of a gene	SO:0001631	Upstream
downstream_gene_variant	A sequence variant located 3' of a gene	SO:0001632	Downstream
TFBS_ablation	A feature ablation whereby the deleted region includes a transcription factor binding site	SO:0001895	Tfbs ablation
TFBS_amplification	A feature amplification of a region containing a transcription factor binding site	SO:0001892	Tfbs amplification
TF_binding_site_variant	A sequence variant located within a transcription factor binding site	SO:0001782	Regulatory region
regulatory_region_variant	A sequence variant located within a regulatory region	SO:0001566	
regulatory_region_ablation	A feature ablation whereby the deleted region includes a regulatory region	SO:0001894	Regulatory region ablation
regulatory_region_amplification	A feature amplification of a region containing a regulatory region	SO:0001891	Regulatory region amplification
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence	SO:0001907	Feature elongation
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	SO:0001906	Feature truncation
intergenic_variant	A sequence variant located in the intergenic region, between genes	SO:0001628	Intergenic

http://www.ensembl.org/info/docs/variation/predicted_data.html

<http://training.ensembl.org/events>



Genes and Regulation

rs34424986 SNP

Most severe consequence

Alleles

Location

Co-located variants

Evidence status ⓘ

Clinical significance ⓘ

HGVS names

Synonyms

Genotyping chips

Original source

About this variant

Description from SNPedia

Missense variant | See all predicted consequences

G/A | Ancestral: G | MAF: < 0.01 (A) | Highest population MAF: < 0.01

Chromosome 6:161785820 (forward strand) | View in location tab

COSMIC COSM1722445, COSM1722444 ; HGMD-PUBLIC CM991007



This variant has 21 HGVS names - Show +

This variant has 4 synonyms - Hide □

- ClinVar RCV000272835, RCV000007466
- LSDB 11251
- Uniprot VAR_019752

This variant has assays on: Illumina_ExomeChip, Illumina_ImmunoChip

Variants (including SNPs and indels) imported from dbSNP (release 149) | View in dbSNP

This variant overlaps 7 transcripts, has 2505 sample genotypes, is associated with 5 phenotypes and is mentioned in 3 citations.

c.823C>T (p.Arg275Trp)

Concise summary

Explore this variant ⓘ



<http://training.ensembl.org/events>



Genes and Regulation

Tables show the predicted impact of a variant on all of the transcripts it overlaps

Genes and regulation ?

Gene and Transcript consequences

Gene	Transcript (strand)	Allele (Tr. allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	AA	Codons	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor	Detail
ENSG0000185345 HGNC: PRKN	ENST0000338468.7 (-) biotype: protein_coding	A (T)	missense variant	701 (out of 1276)	250 (out of 825)	84 (out of 274)	R/W	CGG/TGG	0	1	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000366892.5 (-) biotype: protein_coding	A (T)	missense variant	919 (out of 1505)	823 (out of 1107)	275 (out of 368)	R/W	CGG/TGG	0	0.999	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000366894.5 (-) biotype: protein_coding	A (T)	missense variant	582 (out of 1157)	250 (out of 825)	84 (out of 274)	R/W	CGG/TGG	0	1	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000366896.5 (-) biotype: protein_coding	A (T)	missense variant	479 (out of 2514)	376 (out of 951)	126 (out of 316)	R/W	CGG/TGG	0	0.993	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000366897.5 (-) biotype: protein_coding	A (T)	missense variant	842 (out of 2877)	739 (out of 1314)	247 (out of 437)	R/W	CGG/TGG	0	1	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000366898.5 (-) biotype: protein_coding	A (T)	missense variant	926 (out of 4180)	823 (out of 1398)	275 (out of 465)	R/W	CGG/TGG	0	1	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000479615.5 (-) biotype: nonsense-mediated_decay	A (T)	missense variant NMD transcript variant	659 (out of 904)	586 (out of 657)	196 (out of 218)	R/W	CGG/TGG	0	1	25	0.747	0.818	0.717	Show
ENSG0000185345 HGNC: PRKN	ENST0000338468.7 (-) biotype: protein_coding	T (A)	synonymous variant	701 (out of 1276)	250 (out of 825)	84 (out of 274)	R	CGG/AGG	-	-	-	-	-	-	Show
ENSG0000185345 HGNC: PRKN	ENST0000366892.5 (-) biotype: protein_coding	T (A)	synonymous variant	919 (out of 1505)	823 (out of 1107)	275 (out of 368)	R	CGG/AGG	-	-	-	-	-	-	Show
ENSG0000185345 HGNC: PRKN	ENST0000366894.5 (-) biotype: protein_coding	T (A)	synonymous variant	582 (out of 1157)	250 (out of 825)	84 (out of 274)	R	CGG/AGG	-	-	-	-	-	-	Show
ENSG0000185345 HGNC: PRKN	ENST0000366896.5 (-) biotype: protein_coding	T (A)	synonymous variant	479 (out of 2514)	376 (out of 951)	126 (out of 316)	R	CGG/AGG	-	-	-	-	-	-	Show



The SIFT and PolyPhen2 packages predict how likely a variant is to be damaging to a protein. Red indicates damaging; green indicates tolerated; blue indicates a low quality prediction

<http://training.ensembl.org/events>



Ensembl variation process



Where can I find variation data?

- Website www.ensembl.org
- [Variant Effect Predictor \(VEP\)](#)
- [BioMart](#) (custom download)
- [FTP](#) site
- Programmatically:
 - [Perl API](#) (including [VEP](#))
 - [REST API](#)

http://www.ensembl.org/info/genome/variation/tools/data_access.html

<http://training.ensembl.org/events>



Summary

The Ensembl genome browser displays:

- imported variation data, including phenotype associations, allele frequencies and literature citations from a variety of different sources
- additional variant annotation, including functional consequence predictions
- this data can be accessed through the web-interface, BioMart, Perl API, REST API and the FTP site

<http://training.ensembl.org/events>



Ensembl Acknowledgements

The Entire Ensembl Team

Fiona Cunningham, Premanand Achuthan, Wasiu Akanni, James Allen, M Ridwan Amode, Irina M Armean, Ruth Bennett, Jyothish Bhai, Konstantinos Billis, Sanjay Boddu, Carla Cummins, Claire Davidson, Kamalkumar Jayantilal Dodiya, Astrid Gall, Carlos García Girón, Laurent Gil, Tiago Grego, Leanne Haggerty, Erin Haskell, Thibaut Hourlier, Osagie G Izuogu, Sophie H Janacek, Thomas Juettemann, Mike Kay, Matthew R Laird, Ilias Lavidas, Zhicheng Liu, Jane E Loveland, José C Marugán, Thomas Maurel, Aoife C McMahon, Benjamin Moore, Joannella Morales, Jonathan M Mudge, Michael Nuhn, Denye Ogeh, Anne Parker, Andrew Parton, Mateus Patrício, Ahamed Imran Abdul Salam, Bianca M Schmitt, Helen Schuilenburg, Dan Sheppard, Helen Sparrow, Eloise Stapleton, Marek Szuba, Kieron Taylor, Glen Threadgold, Anja Thormann, Alessandro Vullo, Brandon Walts, Andrea Winterbottom, Amonida Zadissa, Marc Chakiachvili, Adam Frankish, Sarah E Hunt, Myrto Kostadima, Nick Langridge, Fergal J Martin, Matthieu Muffato, Emily Perry, Magali Ruffier, Daniel M Staines, Stephen J Trevanion, Bronwen L Aken, Andrew D Yates, Daniel R Zerbino, Paul Flicek¹

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Funding



Co-funded by
the European
Union

Training materials

- Ensembl training materials are protected by a CC BY license
- <http://creativecommons.org/licenses/by/4.0/>
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers
- <http://www.ensembl.org/info/about/publications.html>



<http://training.ensembl.org/events>

