

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: the Ensembl browser and its possibilities

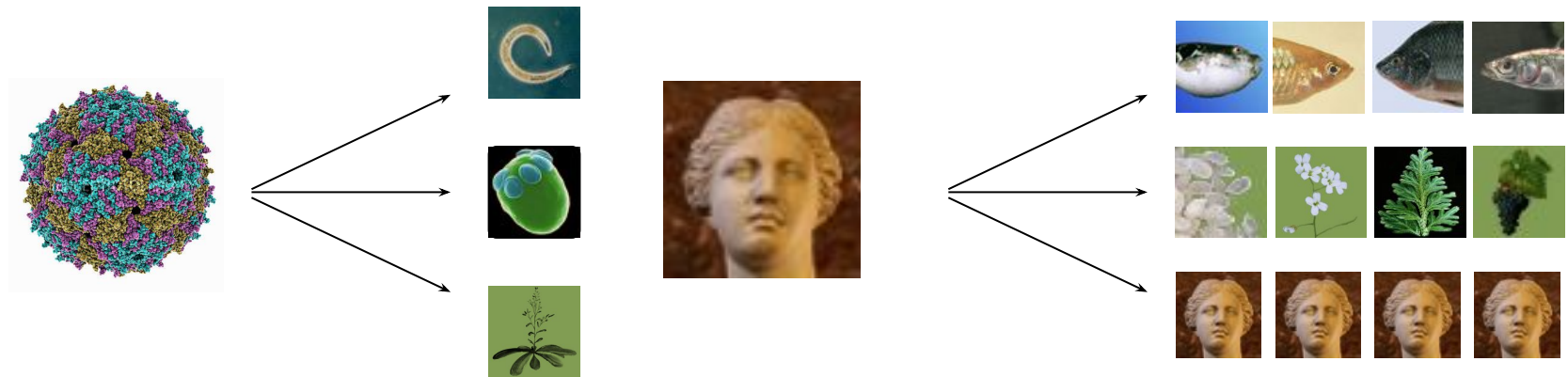
Benjamin Moore
Ensembl Outreach Officer

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)

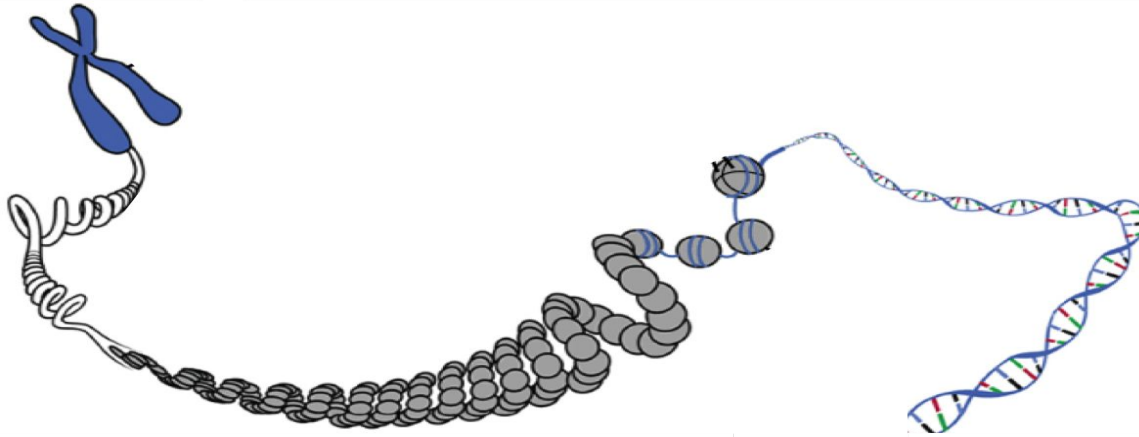


<http://training.ensembl.org/events>

Why do we need genome browsers?

CGGCCTTTGGGCTCCGCCTTCAGCTCAAGACTTAACTTCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTTCTTGGAACACGATCAO
TTTAACGGAATATTGCTGTTTTGGGAAGTGTTTTACAGCTGCTGGGCACGCTGTATTTGCCTTACTTAAGCCCCTGGTAATTGCTGTATTG
CGAAGACATGCTGATGGGAATTACCAGGCGGCGTTGGTCTCTAACTGGAGCCCTCTGTCCCCACTAGCCACGCGTCACTGGTTAGCGTGATT
GAAACTAAATCGTATGAAAATCCTCTTCTCTAGTCGCACTAGCCACGTTTCGAGTGCTTAATGTGGCTAGTGGCACCGGTTTGGACAGCACA
GCTGTAAAATGTTCCCATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGACTGAATTAGAATTCAAACAAATTTTCCAGCGCTTCTGA
GTTTTACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTTTTGGTCTTCTGTTTTTGCAGACTTATTTACCAAGCATTGGAGGA
ATATCGTAGGTAAAAATGCCTATTGGATCCAAAGAGAGGCCAACATTTTTTGAATTTTTTAAGACACGCTGCAACAAAGCAGGTATTGACAA
ATTTTATATAACTTTATAAATTACACCGAGAAAGTGTTTTCTAAAAAATGCTTGCTAAAAACCCAGTACGTCACAGTGTTGCTTAGAACCAT
AAACTGTTCCCTTATGTGTGTATAAATCCAGTTAACAACATAATCATCGTTTGCAGGTTAACCACATGATAAATATAGAACGTCTAGTGGATA
AAGAGGAAACTGGCCCCTTGACTAGCAGTAGGAACAATTACTAACAATCAGAAGCATTAAATGTTACTTTATGGCAGAAGTTGTCCAACTTT
TTGGTTTCAGTACTCCTTATACTCTTAAAAATGATCTAGGACCCCCGGAGTGCTTTTGTATTATGTAGCTTACCATATTAGAAATTTAAACT
AAGAATTTAAGGCTGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACTTGAGGCCAGAAGTTTGA
GACCAGCCTGGCCAACATGGTGAAACCCTATCTCTACTAAAAATACAAAAAATGTGCTGCGTGTGGTGGTGCCTGTAATCCCAGCTAC
ACGGGAGGTGGAGGCAGGAGAATCGCTTGAACCCTGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTCTAGCCTGGGCCACA
TAGCATGACTCTGTCTCAAACAAACAAACAAACAAAACTAAGAATTTAAAGTTAATTTACTTAAAAATAATGAAAGCTAACCATTGCA
TATTATCACAAACATTCTTAGGAAAAATAACTTTTTTGAACAAGTGAGTGGAATAGTTTTTACATTTTTTGCAGTTCTCTTTAATGTCTGGCT
AAATAGAGATAGCTGGATTCACCTTATCTGTGTCTAATCTGTTATTTTTGGTAGAAGTATGTGAAAAAAATTAACCTCACGTTGAAAAAAGGA
ATATTTTAATAGTTTTTCAGTTACTTTTTTGGTATTTTTTCCCTTGTAATTTTGCATAGATTTTTTCAAAGATCTAATAGATATACCATAGGTCTTT
CCATGTGCGAACATCATGCAGTGATTATTTTGAAGATAGTGGTGTCTGAATTATACAAAGTTTCCAAATATTGATAAATTGCATTAAACTA
TTTTAAAAATCTCATTCAATTAATACCACCATGGATGTCAGAAAAGTCTTTTTAAGATTGGGTAGAAATGAGCCACTGGAAATTCTAATTTTCA
TTTGAAAGTTCACATTTTGTCAATTGACAACAACTGTTTTTCCCTTGCAAGCAACAAGATCACTTCATTGATTTGTGAGAAAATGTCTACCAAAT
TATTTAAGTTGAAATAACTTTGTGAGCTGTTCTTTCAAGTAAAAATGACTTTTCATTGAAAAAATTGCTTGTTCAGATCACAGCTCAACATG
AGTGCTTTTCTAGGCAGTATTGTACTTCAGTATGCAGAAGTGCTTTATGTATGCTTCCTATTTTGTGAGAGATTATTAAGAAGTGCTAAA
GCATTGAGCTTCGAAATTAATTTTTTACTGCTTCATTAGGACATTCTTACATTAACTGGCATTATTATTACTATTATTTTAAACAAGGACAC
TCAGTGGTAAGGAATATAATGGCTACTAGTATTAGTTTGGTGCCACTGCCATAACTCATGCAAATGTGCCAGCAGTTTTACCCAGCATCATC
TTTGCAGTGTGATACAAATGTCAACATCATGAAAAAGGGTTGAAAAAAGGAATATTTTAATAGTTTTTCAAGTTACTTTATGACTGTTAGCTA
<http://training.ensembl.org/events>

Ensembl- unlocking the code

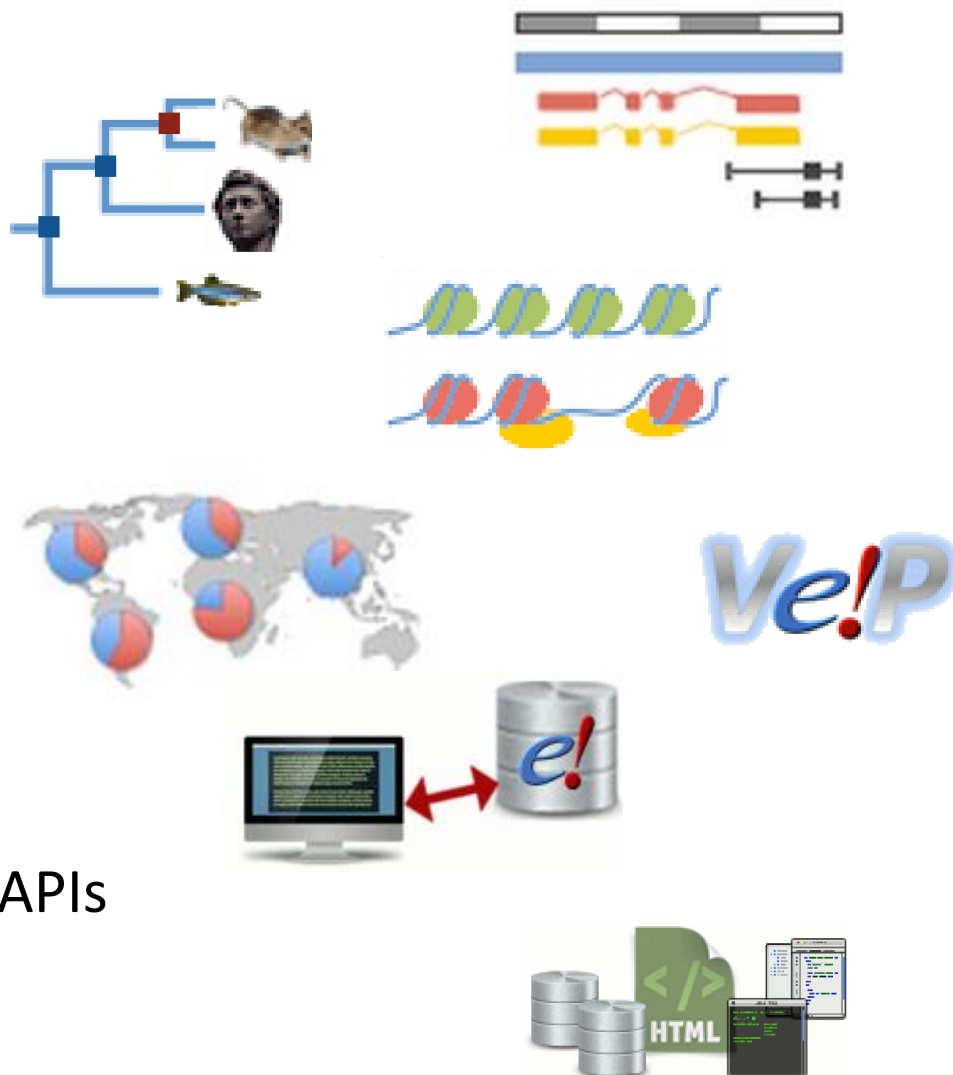


- Genomic assemblies - automated gene annotation
- Variation - Small and large scale sequence variation with phenotype associations
- Comparative Genomics - Whole genome alignments, gene trees
- Regulation - Potential promoters and enhancers, DNA methylation

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

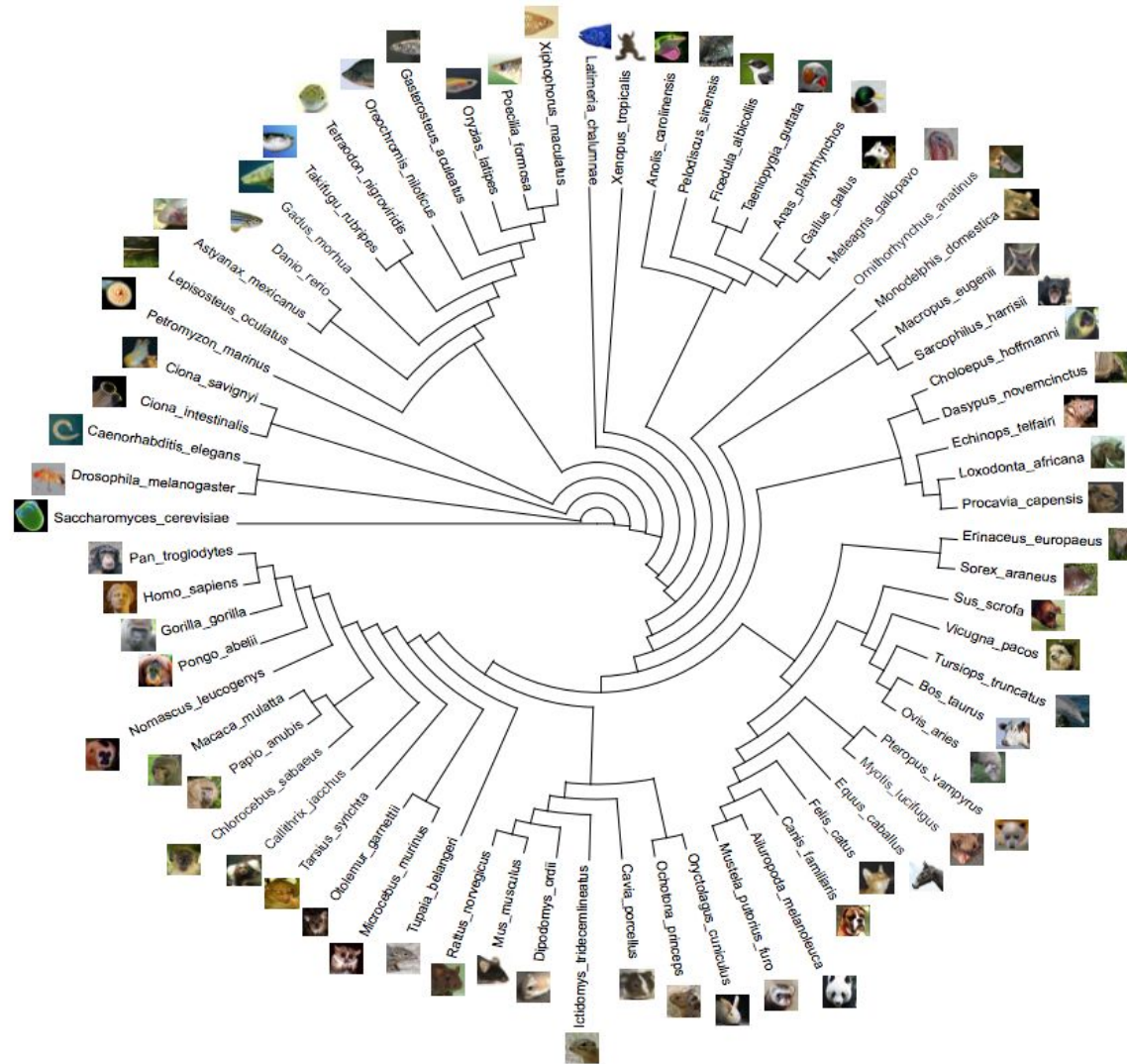
Ensembl Features

- 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- access to 100+ genomes



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

Ensembl Genomes- expanding Ensembl



www.ensembl.org

- Vertebrates



- Other representative species

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

Ensembl Genomes- expanding Ensembl



www.ensembl.org

- Vertebrates

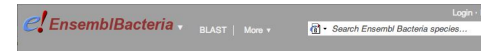


- Other representative species

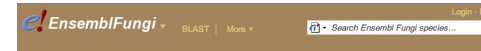
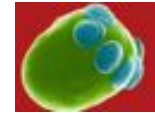


www.ensemblgenomes.org

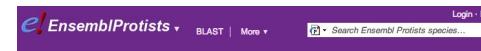
- Bacteria



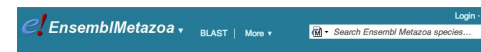
- Fungi



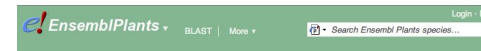
- Protists



- Metazoa



- Plants



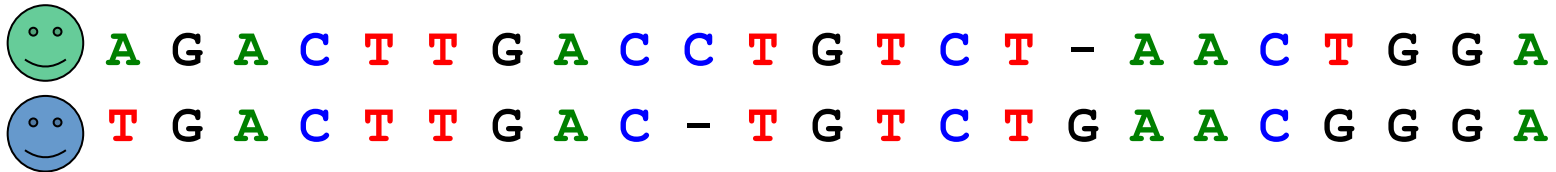
<http://training.ensembl.org/events>



Variation types

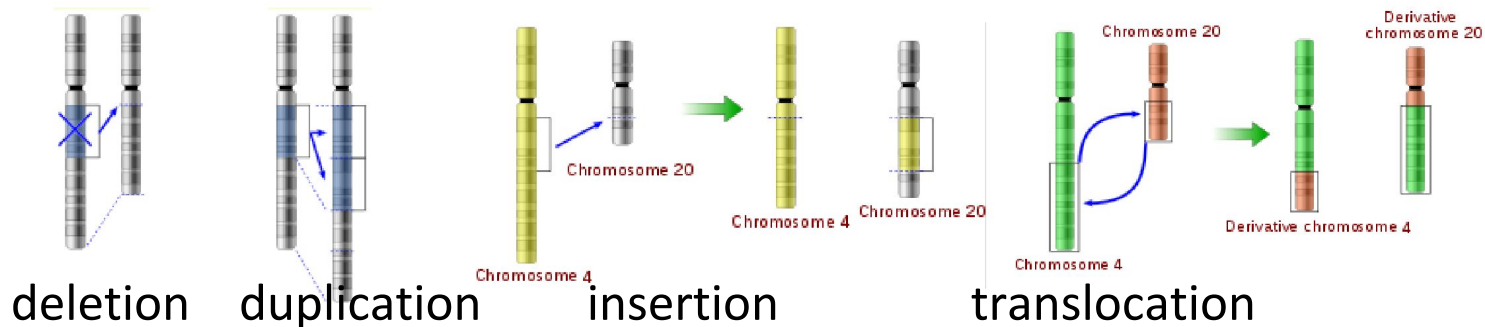
1) Small scale in one or few nucleotides of a gene

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



2) Large scale in chromosomal structure (structural variation)

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



22 species with variation data

 Cat <i>Felis catus</i>	3.6 million+	 Opossum <i>Monodelphis domestica</i>	1.1 million+
 Chicken <i>Gallus gallus</i>	21 million+	 Orangutan <i>Pongo abelii</i>	10 million+
 Chimpanzee <i>Pan troglodytes</i>	1.6 million+	 Pig <i>Sus scrofa</i>	60 million+
 Cow <i>Bos taurus</i>	100 million+	 Platypus <i>Ornithorhynchus anatinus</i>	1.3 million+
 Dog <i>Canis familiaris</i>	5.9 million+	 Rat <i>Rattus norvegicus</i>	5 million+
 Fruitfly <i>Drosophila melanogaster</i>	6.7 million+	 S. cerevisiae <i>Saccharomyces cerevisiae</i>	263,000+
 Gibbon <i>Nomascus leucogenys</i>	1.1 million+	 Sheep <i>Ovis aries</i>	51 million+
 Horse <i>Equus caballus</i>	5 million+	 Tetraodon <i>Tetraodon nigroviridis</i>	902,000+
 Human <i>Homo sapiens</i>	329 million+	 Turkey <i>Meleagris gallopavo</i>	9,000+
 Macaque <i>Macaca mulatta</i>	3 million+	 Zebra Finch <i>Taeniopygia guttata</i>	1.7 million+
 Mouse <i>Mus musculus</i>	80 million+	 Zebrafish <i>Danio rerio</i>	17 million+

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

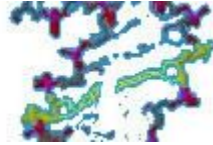
22 species with variation data

 Cat <i>Felis catus</i> 3.6 million+	 Opossum <i>Monodelphis domestica</i> 1.1 million+
 Chicken <i>Gallus gallus</i> 21 million+	 Orangutan <i>Pongo abelii</i> 10 million+
 Chimpanzee <i>Pan troglodytes</i> 1.6 million+	 Pig <i>Sus scrofa</i> 60 million+
 Cow <i>Bos taurus</i> 100 million+	 Platypus <i>Ornithorhynchus anatinus</i> 1.3 million+
 Dog <i>Canis familiaris</i> 5.9 million+	 Rat <i>Rattus norvegicus</i> 5 million+
 Fruitfly <i>Drosophila melanogaster</i> 6.7 million+	 S. cerevisiae <i>Saccharomyces cerevisiae</i> 263,000+
 Gibbon <i>Nomascus leucogenys</i> 1.1 million+	 Sheep <i>Ovis aries</i> 51 million+
 Horse <i>Equus caballus</i> 5 million+	 Tetraodon <i>Tetraodon nigroviridis</i> 902,000+
 Human <i>Homo sapiens</i> 650 million+ 320 million+	 Turkey <i>Meleagris gallopavo</i> 9,000+
 Macaque <i>Macaca mulatta</i> 3 million+	 Zebra Finch <i>Taeniopygia guttata</i> 1.7 million+
 Mouse <i>Mus musculus</i> 80 million+	 Zebrafish <i>Danio rerio</i> 17 million+

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

Variation sources

dbSNP
Short Genetic Variations



orphanet

DECIPHER v5.1
GRCh37

OMIM[®]

DGVar^{archive}

PubMed



UniProt

European
genome-phenome
archive



GANT



GEFS

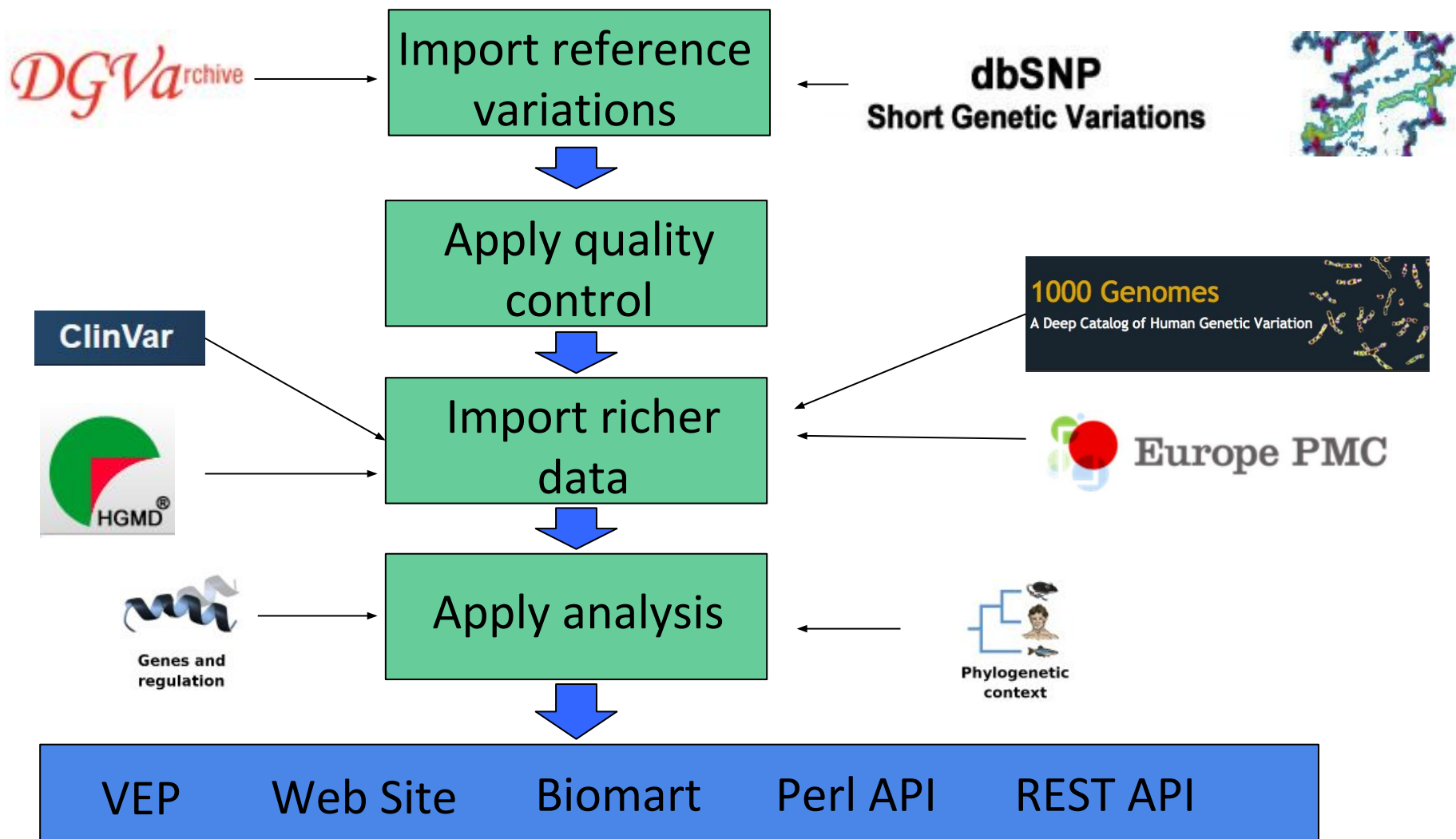
Genetics

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

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



The Ensembl variation process



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

Quality Control

Icon	Name	Description
	Multiple observations	The variant has multiple independent dbSNP submissions, i.e. submissions with a different source
	Frequency	The variant is reported to be polymorphic in at least one sample
	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)
	ESP	The variant was discovered in the Exome Sequencing Project (human only).

We summarise the information supporting each dbSNP variant to give a guide to its reliability

We run basic checks and flag suspicious variants

rs14390 SNP

This variation has been flagged

- Flagged as suspect by dbSNP
- Variation maps to 2 genomic locations

Select a location:

Original source

Alleles

Location

Evidence status

Variants (including SNPs and indels) imported from dbSNP (release 138)

C/T | Ancestral: T | Ambiguity code: Y

This variation maps to 2 genomic locations; None selected



QC Type	Reported failure reason
Mapping checks	Variant does not map to the genome
	Variant maps to more than 1 location
	Mapped position is not compatible with reported alleles
Checks on the alleles of refSNPs	None of the variant alleles match the reference allele
	Loci with no observed variant alleles in dbSNP
	Alleles contain ambiguity codes
Checks on the alleles in dbSNP submissions	Alleles contain non-nucleotide characters
	Additional submitted allele data from dbSNP does not agree with the dbSNP refSNP alleles
External failure classification	Flagged as suspect by dbSNP

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

Finding your variant

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | More ▾

Login/Register

Search: for

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.

Find a Data Display

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

TABLE
HEATMAP
SEQUENCE
PIE CHART

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](#)

Favourite genomes

 **Human**
GRCh38.p10

 **Mouse**
GRCm38.p5

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
- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

rs1800566 SNP

Most severe consequence
Alleles
Location
Co-located variant
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: **0.29** (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 **G** at this location denotes the NQO1*2 allele.... [Show](#) ⓘ

Explore this variant ⓘ

Genomic context (8)
Genes and regulation
Flanking sequence
Population genetics (115)
Phenotype data (6)
Sample genotypes (4674)
Linkage disequilibrium
Phylogenetic context
Citations (159)

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

Menu of pages, available on all pages

Icons available (grey available)

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

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:PJL	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.

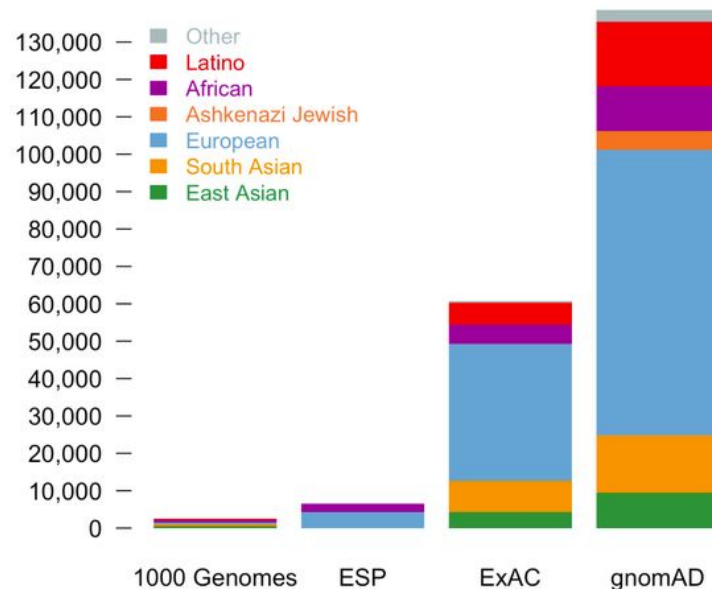
<http://training.ensembl.org/events>

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
- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

rs1800566 SNP

Most severe consequence

Alleles

Location

Co-located variant

Evidence status

Clinical significance

HGVS names

Synonyms

Genotyping chips

Original source

About this variant

Description from SNpedia

Explore this variant

Concise summary

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

Chromosome 16:69,710,742 (forward strand) | Ref: 16:69,710,742 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](#)

Genomic context

Genes and regulation

Flanking sequence

Population genetics

Phenotype data

Sample genotypes

Linkage disequilibrium

Phylogenetic context

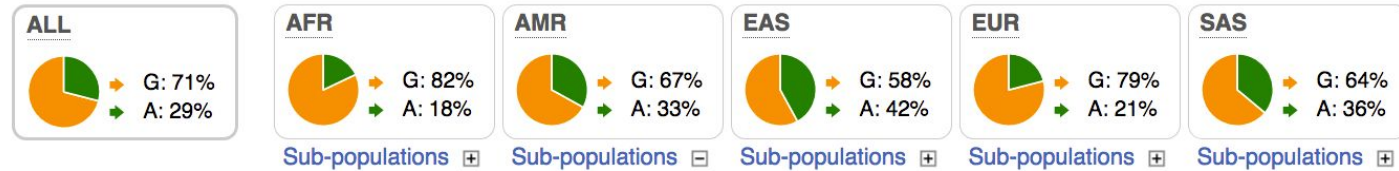
Citations

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

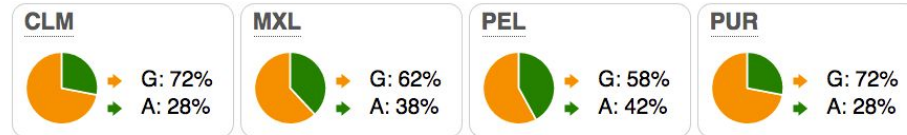
Allele Frequency Distributions

Population genetics ?

1000 Genomes Project Phase 3 allele frequencies



AMR sub-populations



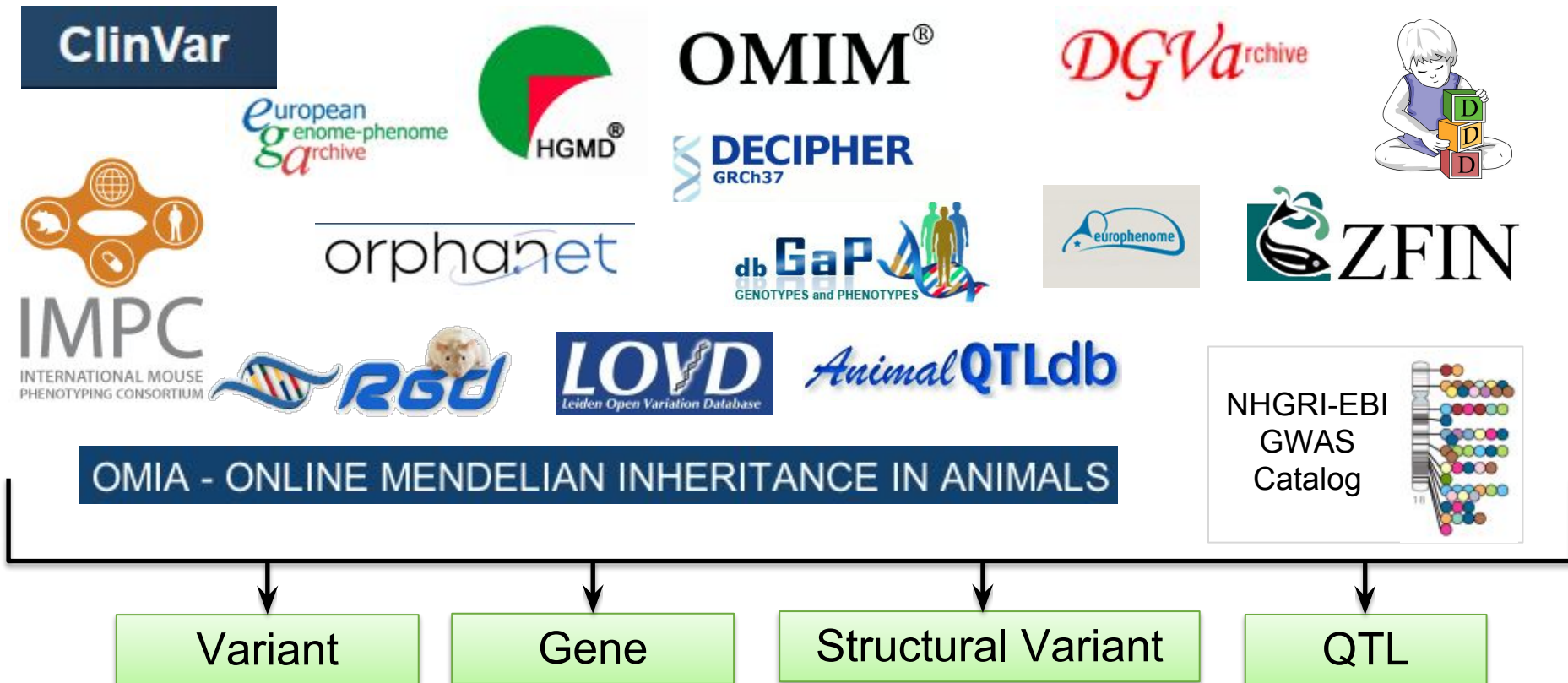
gnomAD exomes (9)

Show/hide columns		Filter	
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>

Phenotype and Disease Data



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

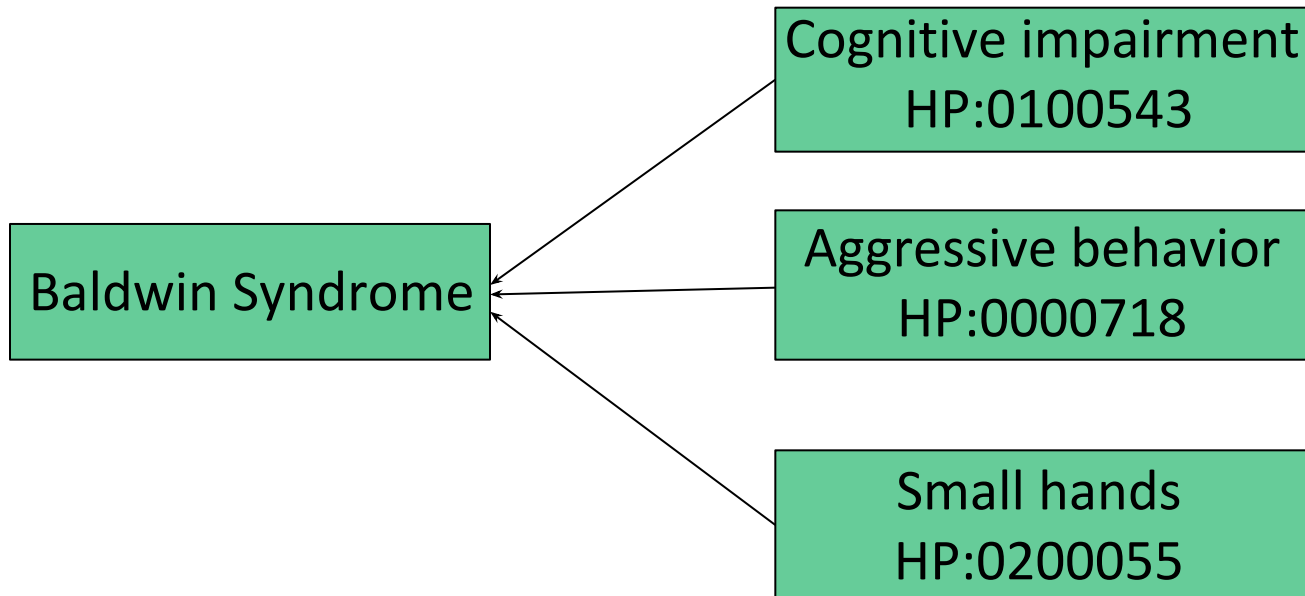
<http://training.ensembl.org/events>

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.

Disease

Phenotypes



Hand size can also be considered a **trait**

The Naming Problem

Different sources describe the phenotypes/ diseases in different ways

- We need to normalise the descriptions for easy querying
- Also need to respect the original data

Phenotype Data ?

Significant association(s)

rs2470893 SNP

Show	All	entries	Show/hide columns (3 hidden)	Filter	
Phenotype, disease and trait	Source(s)	Study	Reported gene(s)	Associated allele	Statistics
Caffeine consumption	NHGRI-EBI GWAS catalog	PMID:21490707	EDC3 , CYP1A1 , CYP1A2 , LMAN1L , CSK	T	p-value: 5.00e-14 beta coefficient: 0.12 mg/day increase
Coffee consumption	NHGRI-EBI GWAS catalog	PMID:21876539	CYP1A1 , CYP1A2	T	p-value: 2.00e-11 beta coefficient: 0.0675 unit increase
Coffee consumption	NHGRI-EBI GWAS catalog	PMID:25288136	CYP1A1 , CYP1A2	T	p-value: 5.00e-19 beta coefficient: 0.2 unit increase

Features are often reported as associated with a disease sub-type making it complex to extract all loci of interest

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

The Solution

Use phenotype and disease ontologies to organise descriptions and allow query and display by synonym and child/parent term.

OLS > Experimental Factor Ontology **EFO** > **EFO:0004330** 

coffee consumption

 http://www.ebi.ac.uk/efo/EFO_0004330 

Behaviors associated with the ingesting of coffee

Synonyms: caffeine consumption

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
- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

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

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

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

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

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

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

Description from SNpedia 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](#)

Very concise summary

Explore this variant ?

8

115

6

4674

159

Genomic context

Genes and regulation

Flanking sequence

Population genetics

Phenotype data

Sample genotypes

Linkage disequilibrium

Phylogenetic context

Citations

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

Phenotype and Disease Data

Phenotype Data ?

Significant association(s)

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	

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

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
- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

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: CM950861](#)

Evidence status

Clinical significance

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

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

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

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

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

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

Description from SNPedia 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>

Citation Data

Citations

rs1800566 is mentioned in the following publications

Show All entries		Show/hide columns		Filter		
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, Alcaide-Cubero S, Ayuso-Fernández J, Benito-León J, Benito-Pisa D, Benito-Pisa P, Ortega-García-Albea E, Plaza-Nieto JF, Jiménez-Jiménez FJ.	PMC4751624		

Link to
full text

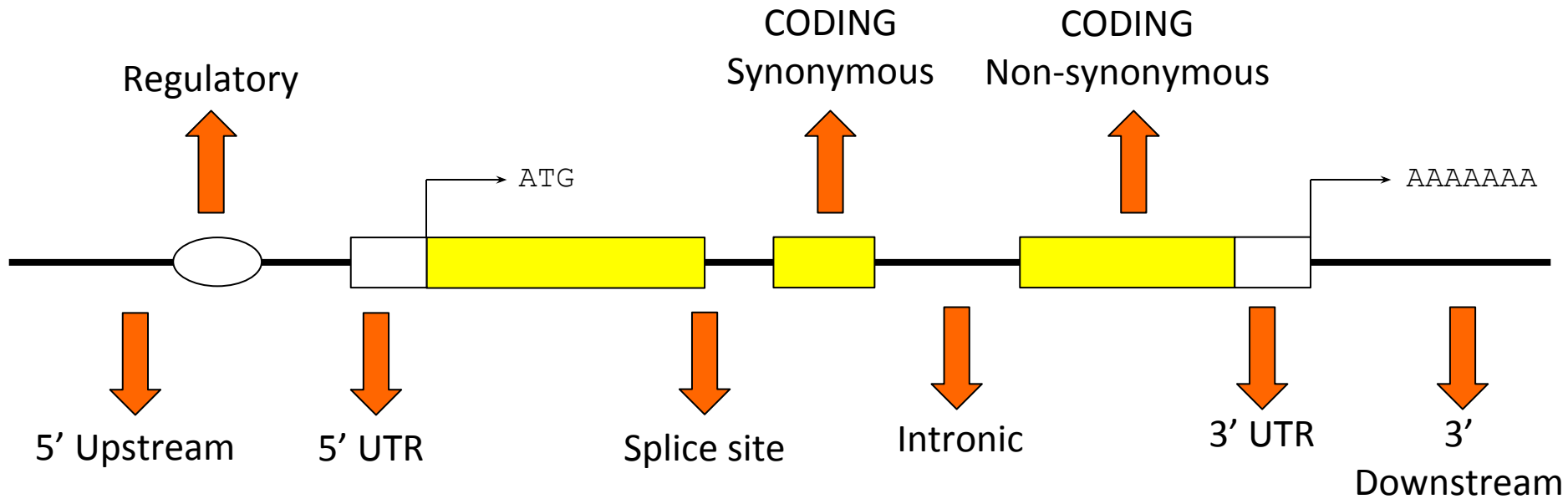
<http://training.ensembl.org/events>

Variant Annotation

We analyse the imported variation data utilising other Ensembl resources. Annotation provided includes:

- Comparison to Ensembl gene structures to predict transcript/ protein consequences
- Comparison to Ensembl regulatory features to find co-location
- Comparison to other species to predict ancestral allele, show phylogenetic context
- Linkage disequilibrium calculations for the 1000 Genomes populations

Variant consequences



We calculated the predicted consequences of variants on the Ensembl transcript set and assign Sequence Ontology terms

CAGCATCAACTGCTGCAGCAACATCAGCAGCAGCATCAGCAATCAGCATCAACTGCTGCAGCAACATCAGCAGCAGCATCAGCA



<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 SNPeedia

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 [CM991607](#)



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

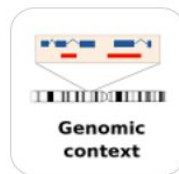
Variants (including SNPs and indels) imported from dbSNP (release 149) | [View in dbSNP](#) ⓘ

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 ⓘ



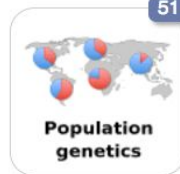
Genomic
context



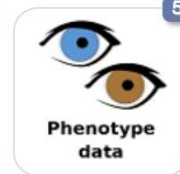
Genes and
regulation



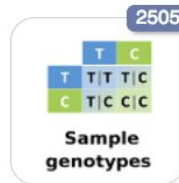
Flanking
sequence



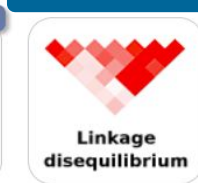
Population
genetics



Phenotype
data



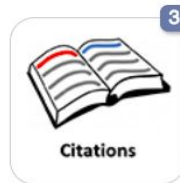
Sample
genotypes



Linkage
disequilibrium



Phylogenetic
context



Citations

<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

Show All entries Show/hide columns Filter											
Gene	Transcript (strand)	Allele (transcript allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	Amino acid	Codons	SIFT	PolyPhen	Detail
ENSG00000185345 HGNC: PRKN	ENST00000338468.7 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000366892.5 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000366894.5 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000366896.5 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000366897.5 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000366898.5 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000479615.5 (-) 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	Show
ENSG00000185345 HGNC: PRKN	ENST00000338468.7 (-) biotype: protein_coding	T (A)	synonymous variant	701 (out of 1276)	250 (out of 825)	84 (out of 274)	R	CGG/AGG	-	-	Show
ENSG00000185345 HGNC: PRKN	ENST00000366892.5 (-) biotype: protein_coding	T (A)	synonymous variant	919 (out of 1505)	823 (out of 1107)	275 (out of 368)	R	CGG/AGG	-	-	Show
ENSG00000185345 HGNC: PRKN	ENST00000366894.5 (-) biotype: protein_coding	T (A)	synonymous variant	582 (out of 1157)	250 (out of 825)	84 (out of 274)	R	CGG/AGG	-	-	Show
ENSG00000185345 HGNC: PRKN	ENST00000366896.5 (-) 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>

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

Daniel R. Zerbino¹, Premanand Achuthan¹, Wasiu Akanni¹, M. Ridwan Amode¹, Daniel Barrell^{1,2}, Jyothish Bhai¹, Konstantinos Billis¹, Carla Cummins¹, Astrid Gall¹, Carlos García Giroñ¹, Laurent Gil¹, Leo Gordon¹, Leanne Haggerty¹, Erin Haskell¹, Thibaut Hourlier¹, Osagie G. Izuogu¹, Sophie H. Janacek¹, Thomas Juettemann¹, Jimmy Kiang To¹, Matthew R. Laird¹, Ilias Lavidas¹, Zhicheng Liu¹, Jane E. Loveland¹, Thomas Maurel¹, William McLaren¹, Benjamin Moore¹, Jonathan Mudge¹, Daniel N. Murphy¹, Victoria Newman¹, Michael Nuhn¹, Denye Ogeh¹, Chuang Kee Ong¹, Anne Parker¹, Mateus Patricio¹, Harpreet Singh Riat¹, Helen Schuilenburg¹, Dan Sheppard¹, Helen Sparrow¹, Kieron Taylor¹, Anja Thormann¹, Alessandro Vullo¹, Brandon Walts¹, Amonida Zadissa¹, Adam Frankish¹, Sarah E. Hunt¹, Myrto Kostadima¹, Nicholas Langridge¹, Fergal J. Martin¹, Matthieu Muffato¹, Emily Perry¹, Magali Ruffier¹, Dan M. Staines¹, Stephen J. Trevanion¹, Bronwen L. Aken¹, Fiona Cunningham¹, Andrew Yates¹ and Paul Flicek^{1,3}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK, ²Eagle Genomics Ltd., Wellcome Genome Campus, Hinxton, Cambridge CB10 1DR, UK and ³Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, 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>