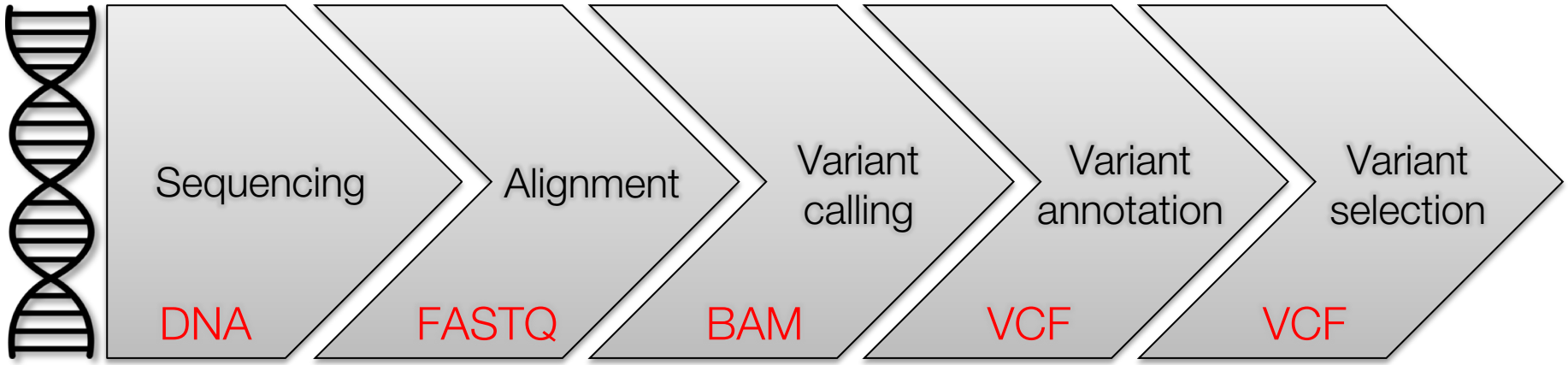
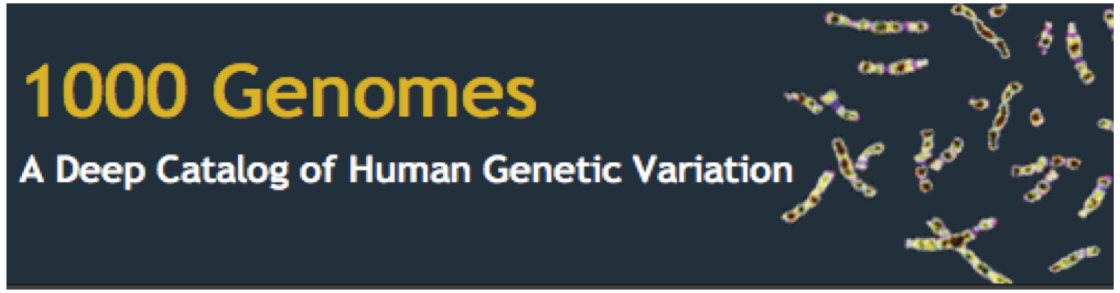
Sequencing workflow



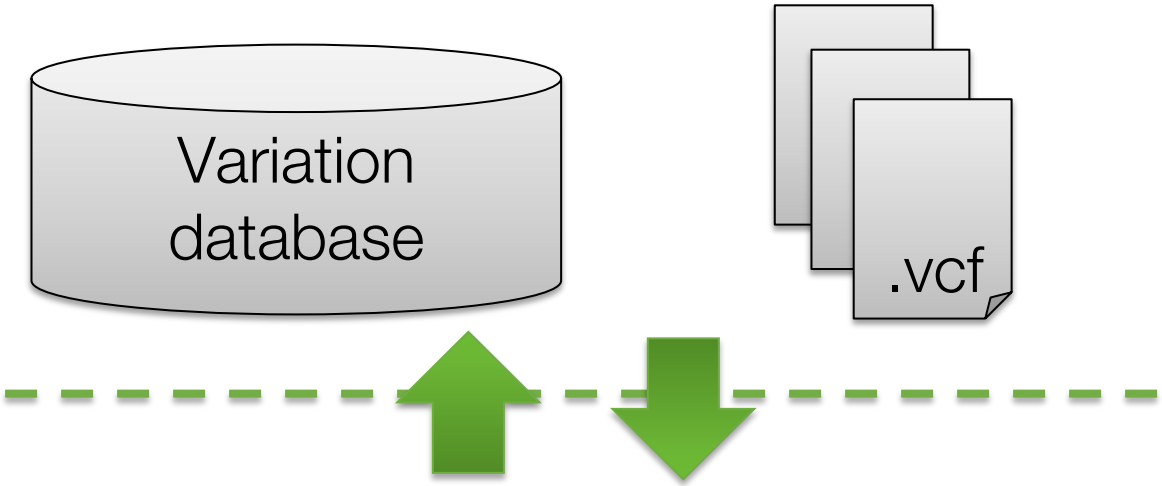


VCF format

1. CHROM – chromosome name
2. POS – reference position
3. ID – unique identifier(s)
4. REF – reference allele
5. ALT – alternate allele(s)
6. QUAL – quality score
7. FILTER – “PASS” or list of failed filters
8. INFO – additional / miscellaneous information
9. FORMAT – genotype specification

```
#CHROM POS     ID        REF ALT    QUAL FILTER INFO              FORMAT NA00001
20     14370   rs6054257 G   A      29   PASS   AF=0.5           GT     0|0
20     17330   .         T   A      3    q10    AF=0.017         GT     1|1
20     1110696 rs6040355 A   G,T    67   PASS   AF=0.333,0.667 GT     0|2
```

VCF backend



Remote @EBI

- Data is stored in databases and VCF files

Local @home

- Use the Ensembl API to
  - compute allele and genotype counts and frequencies
  - Linkage disequilibrium
- calculated on the fly

We have data from VCF files for:

- Human
- Mouse
- Cow
- Sheep
- Goat

The details are specified in a configuration file:

- ensembl-variation/modules/Bio/EnsEMBL/Variation/DBSQL/vcf_config.json

Tell API to look up data from VCF files

```
# Create any adaptor from the variation API
# For example create a variation adaptor
my $va = $registry->get_adaptor('human', 'variation', 'variation');

# Look up data from VCF files and database
$va->db->use_vcf(1)

# Look up data ONLY from VCF files
$va->db->use_vcf(2)
```

Variation

get_all_SampleGenotypes

get_all_PopulationGenotypes

get_all_Alleles

SampleGenotype

- variation
- sample
- genotype_string

PopulationGenotype

- variation
- population
- genotype_string
- frequency

Allele

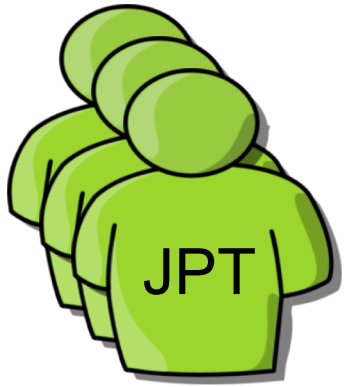
- variation
- population
- allele
- frequency

rs4988235

A|G

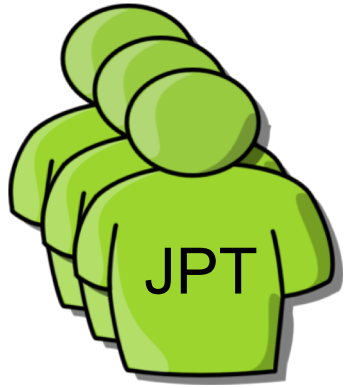
rs4988235

A|A 0.3



rs4988235

A 0.3



Allele

Method	Return type	Example value
`$a->allele()`	String	G
`$a->count()`	Int	34, undef if not defined
`$a->frequency()`	Float	0.15, undef if not defined
`$a->population()`	Population object	Bio::EnsEMBL::Variation::Population, undef if not defined
`$a->variation()`	Variation object	Bio::EnsEMBL::Variation::Variation

SampleGenotype

Method	Return type	Example value
`$sg->genotype()` `@alleles = @{$sg->genotype}`	Genotype List reference	[A, C]
`$sg->genotype_string()`	Genotype String	A|C
`$sg->variation()`	Variation Object	Bio::EnsEMBL::Variation::**Variation**
`$sg->sample()`	Sample Object	Bio::EnsEMBL::Variation::**Sample**
`$sg->sample->individual()`	Individual	Bio::EnsEMBL::Variation::**Individual**

PopulationGenotype

Method	Return type	Example value
`$pg->genotype()` `@alleles = @{$pg->genotype}`	Genotype List reference	[A, C]
`$pg->genotype_string()`	Genotype String	A|C
`$pg->frequency()`		0.5
`$pg->variation()`	Variation Object	Bio::EnsEMBL::Variation::**Variation**
`$pg->population()`	Population Object	Bio::EnsEMBL::Variation::**Population**

Projects	Population names	Allele Frequency	SampleGenotype PopulationGenotype
gnomAD	gnomAD	✓	✗
1000 Genomes Project	1000GENOMES:phase_3	✓	✓
TOPMed	TOPMed	✓	✗
UK10K	TWINSUK ALSPAC	✓	✗

1000 Genomes Project Allele and Genotype Frequencies

rs4988235

EAS	**G: 1.000 (1008)**		**GIG: 1.000 (504)**			
CDX	G: 1.000 (186)		GIG: 1.000 (93)			
CHB	G: 1.000 (206)		GIG: 1.000 (103)			
CHS	G: 1.000 (210)		GIG: 1.000 (105)			
JPT	G: 1.000 (208)		GIG: 1.000 (104)			
KHV	G: 1.000 (198)		GIG: 1.000 (99)			
EUR	**G: 0.492 (495)**	**A: 0.508 (511)**	**GIG: 0.306 (154)**	**AIA: 0.322 (162)**	**AIG: 0.372 (187)**
CEU	G: 0.263 (52)	A: 0.737 (146)	GIG: 0.071 (7)	AIA: 0.545 (54)	AIG: 0.384 (38)
FIN	G: 0.409 (81)	A: 0.591 (117)	GIG: 0.141 (14)	AIA: 0.323 (32)	AIG: 0.535 (53)
GBR	G: 0.280 (51)	A: 0.720 (131)	GIG: 0.099 (9)	AIA: 0.538 (49)	AIG: 0.363 (33)
IBS	G: 0.542 (116)	A: 0.458 (98)	GIG: 0.318 (34)	AIA: 0.234 (25)	AIG: 0.449 (48)
TSI	G: 0.911 (195)	A: 0.089 (19)	GIG: 0.841 (90)	AIA: 0.019 (2)	AIG: 0.140 (15)

Allele

PopulationGenotype

gnomAD Allele Frequencies

rs4988235

Population	Allele: frequency (count)	
gnomADg:**ALL**	G: 0.589 (18204)	A: 0.411 (12712)
gnomADg:**AFR**	G: 0.875 (7620)	A: 0.125 (1092)
gnomADg:**AMR**	G: 0.797 (668)	A: 0.203 (170)
gnomADg:**ASJ**	G: 0.897 (271)	A: 0.103 (31)
gnomADg:**EAS**	G: 0.999 (1621)	A: 0.001 (1)
gnomADg:**FIN**	G: 0.429 (1496)	A: 0.571 (1994)
gnomADg:**NFE**	G: 0.402 (6013)	A: 0.598 (8959)
gnomADg:**OTH**	G: 0.526 (515)	A: 0.474 (465)

gnomAD VCF file does not contain genotypes. Allele frequencies and counts are taken from the INFO field.

Exercise

- Print allele frequencies and allele counts for variant rs4988235
  - Print only results from the following projects: gnomAD, UK10K and TOPMed
    - $population_name =~ /gnomad|topmed|twinsuk|alspac/i
    - Which population reports frequencies for the C allele?
  - Print only results for the following 1000 Genomes populations: 1000GENOMES:phase_3:**FIN**, 1000GENOMES:phase_3:**CHB**, 1000GENOMES:phase_3:**ASW**
    - $population_name eq '1000GENOMES:phase_3:**FIN**'
    - Which population has the highest allele frequency for allele A?

Exercise

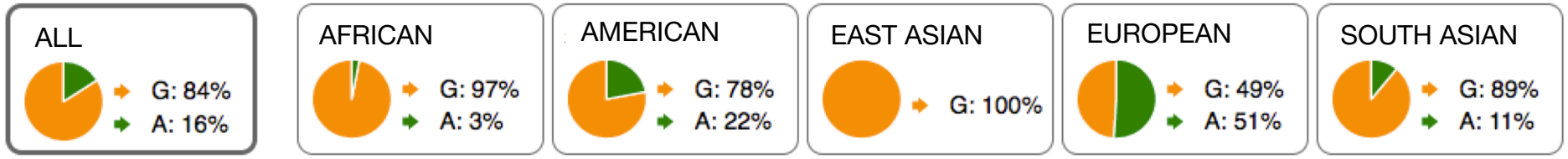
- Find all samples from population 1000GENOMES:phase_3:**ACB** with genotype A|G or G|A for variant rs4988235. Print the sample name and the genotype.
  - `$variation->get_all_SampleGenotypes($population);`

- Print population genotype frequencies and counts observed in 1000GENOMES:phase_3:**ACB** for variant rs4988235.
  - `$variation->get_all_PopulationGenotypes($population)`

Lactose intolerance

- rs4988235: a SNP near the LCT gene controls whether lactase enzyme is turned on or off as a person grows older
- Alleles: G/A

1000 Genomes Project Phase 3 allele frequencies

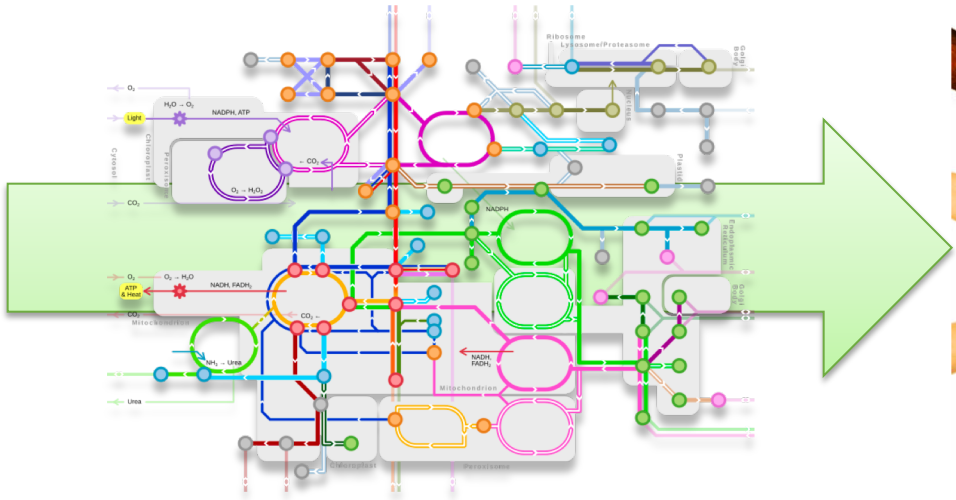


Genotype	What does it mean
G/G	Likely to be lactose intolerant.
G/A	Likely to be tolerant due to lactase persistence.
A/A	Lactose tolerant

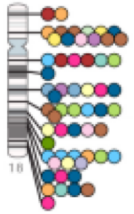
Phenotype data

- Trait
  - Can be described or measured
  - Final product of many molecular and biochemical processes
- Phenotype: Observed state of the trait
- Understand genotype to phenotype link

GIG: 0.318 (34)
AIG: 0.449 (48)
AIA: 0.234 (25)



NHGRI-EBI GWAS Catalog



OMIM®

IMPC
INTERNATIONAL MOUSE PHENOTYPING CONSORTIUM

RGD

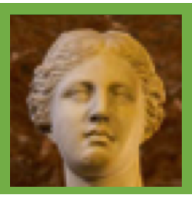
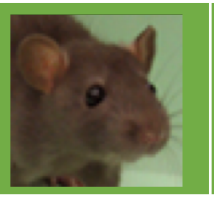
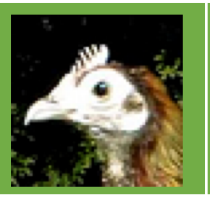
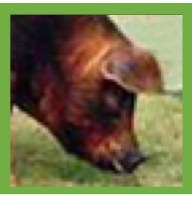
ZFIN

Feature + **Phenotype, Trait, Disease** + **Attributes**

- Genomic Location
  - Gene
  - Variant
  - QTL
  - Structural variant
- Description: as provided by the source
- Clinical significance
- Reported genes
- Most associated risk allele
- P-value
- Inheritance type
- Beta coefficient

PhenotypeFeature

Data available by species

Feature type								
Gene	47736	154,588	64,619	41	132	26	41	50,788
Variant	4,978,589	0	0	0	0	0	0	0
Structural Variant	116,854	0	0	0	0	0	0	0
QTL	0	0	2,266	4,723	100,171	17,588	1,072	0
Sources	17	2	1	3	3	3	3	1

Phenotypes linked to rs2470893

Show All entries | Show/hide columns | Filter

Phenotype, disease and trait	Source(s)	Mapped Terms	Ontology Accessions	Study	Reported gene(s)	Associated allele	Statistics
Caffeine consumption	NHGRI-EBI GWAS catalog	coffee consumption	EFO:0004330	PMID:21490707	EDC3, CYP1A1, CYP1A2, LMAN1L, CSK	T	p-value: $5.00e^{-14}$ beta coefficient: 0.12 mg/day increase
Caffeine metabolism (plasma 1,3,7-trimethylxanthine (caffeine) level)	NHGRI-EBI GWAS catalog	caffeine metabolite measurement	EFO:0007872	PMID:27702941	ARID3B, CLK3, EDC3, CYP1A1	T	p-value: $5.00e^{-18}$ odds ratio: 8.65
Caffeine metabolism (plasma 1,3-dimethylxanthine (theophylline) level)	NHGRI-EBI GWAS catalog	caffeine metabolite measurement	EFO:0007872	PMID:27702941	ARID3B, CLK3, EDC3, CYP1A1	T	p-value: $1.00e^{-7}$ odds ratio: 5.3
Caffeine metabolism (plasma 1,7-dimethylxanthine (paraxanthine) to 1,3,7-trimethylxanthine (caffeine) ratio)	NHGRI-EBI GWAS catalog	caffeine metabolite measurement	EFO:0007872	PMID:27702941	ARID3B, CLK3, EDC3, CYP1A1	T	p-value: $4.00e^{-15}$ odds ratio: 7.85
Coffee consumption	NHGRI-EBI GWAS catalog	coffee consumption, cups of coffee per day measurement	EFO:0004330, EFO:0006782	PMID:21876539	CYP1A1, CYP1A2	T	p-value: $2.00e^{-11}$ beta coefficient: 0.0675 unit increase
Coffee consumption	NHGRI-EBI GWAS catalog	coffee consumption, cups of coffee per day measurement	EFO:0004330, EFO:0006782	PMID:25288136	CYP1A1, CYP1A2	T	p-value: $5.00e^{-19}$ beta coefficient: 0.2 unit increase
Platelet distribution width	NHGRI-EBI GWAS catalog	platelet distribution width	EFO:0007984	PMID:27863252	CYP1A1	T	p-value: $4.00e^{-10}$ beta coefficient: 0.02388056 unit decrease

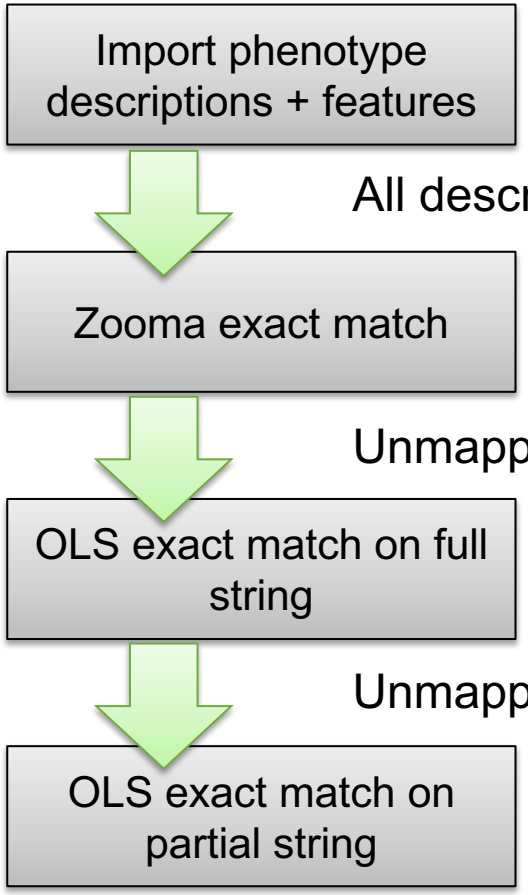
The naming problem

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

The solution

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

Import and mapping process



https://www.ebi.ac.uk/spot/ontology/

Ontology mappings

rs2470893

Show All entries | Show/hide columns | Filter

Phenotype, disease and trait	Source(s)	Mapped Terms	Ontology Accessions	Study	Reported gene(s)	Associated allele	Statistics
Caffeine consumption	NHGRI-EBI GWAS catalog	coffee consumption	EFO:0004330	PMID:21490707	EDC3, CYP1A1, CYP1A2, LMAN1L, CSK	T	p-value: $5.00e^{-14}$ beta coefficient: 0.12 mg/day increase
Caffeine metabolism (plasma 1,3,7-trimethylxanthine (caffeine) level)	NHGRI-EBI GWAS catalog	caffeine metabolite measurement	EFO:0007872	PMID:27702941	ARID3B, CLK3, EDC3, CYP1A1	T	p-value: $5.00e^{-18}$ odds ratio: 8.65
Caffeine metabolism (plasma 1,3-dimethylxanthine (theophylline) level)	NHGRI-EBI GWAS catalog	caffeine metabolite measurement	EFO:0007872	PMID:27702941	ARID3B, CLK3, EDC3, CYP1A1	T	p-value: $1.00e^{-7}$ odds ratio: 5.3
Caffeine metabolism (plasma 1,7-dimethylxanthine (paraxanthine) to 1,3,7-trimethylxanthine (caffeine) ratio)	NHGRI-EBI GWAS catalog	caffeine metabolite measurement	EFO:0007872	PMID:27702941	ARID3B, CLK3, EDC3, CYP1A1	T	p-value: $4.00e^{-15}$ odds ratio: 7.85
Coffee consumption	NHGRI-EBI GWAS catalog	coffee consumption, cups of coffee per day measurement	EFO:0004330, EFO:0006782	PMID:21876539	CYP1A1, CYP1A2	T	p-value: $2.00e^{-11}$ beta coefficient: 0.0675 unit increase
Coffee consumption	NHGRI-EBI GWAS catalog	coffee consumption, cups of coffee per day measurement	EFO:0004330, EFO:0006782	PMID:25288136	CYP1A1, CYP1A2	T	p-value: $5.00e^{-19}$ beta coefficient: 0.2 unit increase
Platelet distribution width	NHGRI-EBI GWAS catalog	platelet distribution width	EFO:0007984	PMID:27863252	CYP1A1	T	p-value: $4.00e^{-10}$ beta coefficient: 0.02388056 unit decrease

http://www.ensembl.org/Homo_sapiens/Variation/Phenotype?r=15:74726608-74727608;v=rs2470893;vdb=variation;vf=1851694

PhenotypeFeature

Method	Return type	Example value
`$pf->type()`	String	Gene, Variation, StructuralVariation, QTL
`$pf->object_id()`	String	Feature name e.g. rs2470893
`$pf->phenotype()`	Phenotype Object	Bio::EnsEMBL::Variation::Phenotype
`$pf->phenotype->description`	String	Caffeine consumption
`$pf->source()`	Source Object	Bio::EnsEMBL::Variation::Source
`$pf->source->name`	String	NHGRI-EBI GWAS catalog
`$pf->associated_gene()`	String	CDKN2BAS,NOTCH2
`$pf->risk_allele()`	String	C
`$pf->p_value()`	Float	6e-07
`$pf->get_all_ontology_accessions`	Arrayref of String	[EFO:0004330, EFO:0006782]
`$pf->seq_region_name()`	String	15, X
`$pf->seq_region_start()` `$pf->seq_region_end()`	Int	74727108 74727108

Add ontologies

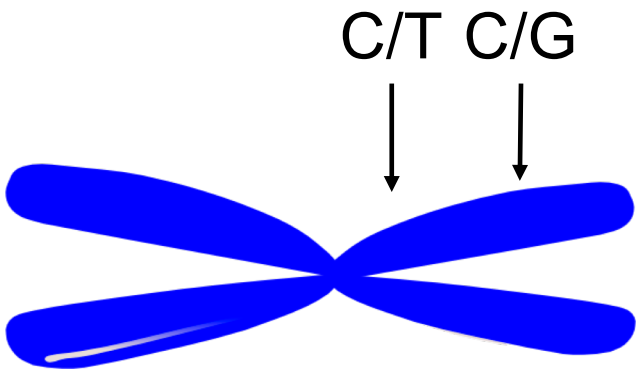
```
# Create a PhenotypeFeature adaptor
# For example create a variation adaptor
my $pfa = $registry->get_adaptor('human', 'variation', 'PhenotypeFeature');

# Add ontology mappings to the output
$pfa->_include_ontology(1);
```

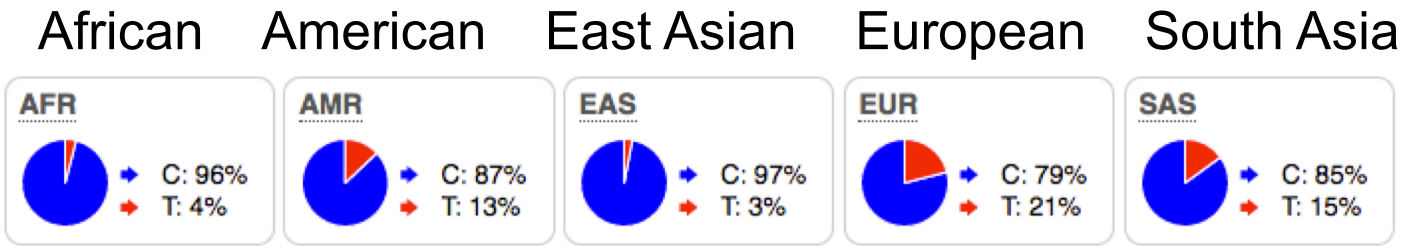
Exercise

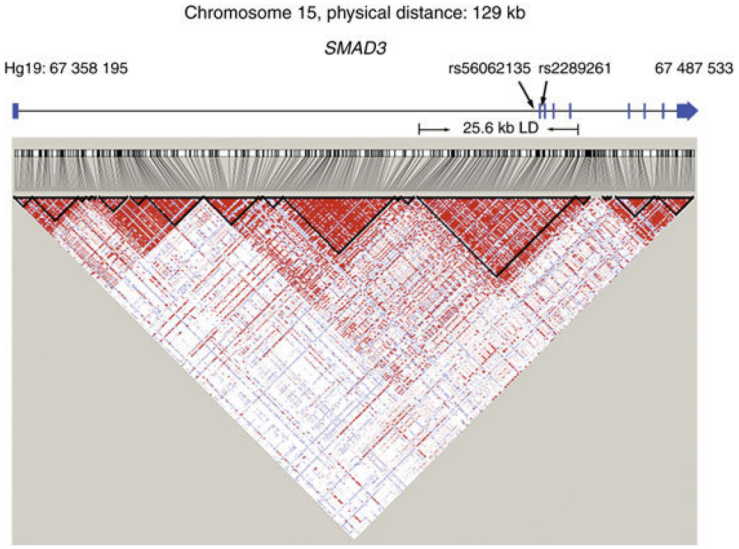
- After reading a publication about the first genome-wide association study of habitual caffeine intake you want to collect all phenotype data that has been linked to rs2470893.
  - Fetch all PhenotypeFeatures for variation rs2470893
  - For each PhenotypeFeature print phenotype description, source name, risk allele, p-value and ontology accessions
- You want to study all features that have been linked to coffee consumption (phenotype accession = EFO:0007872)
  - Fetch all PhenotypeFeatures for phenotype accession EFO:0007872
  - For each PhenotypeFeature print type, object_id, phenotype description and source name. You can use *fetch_all_by_phenotype_accession_source* without providing a source

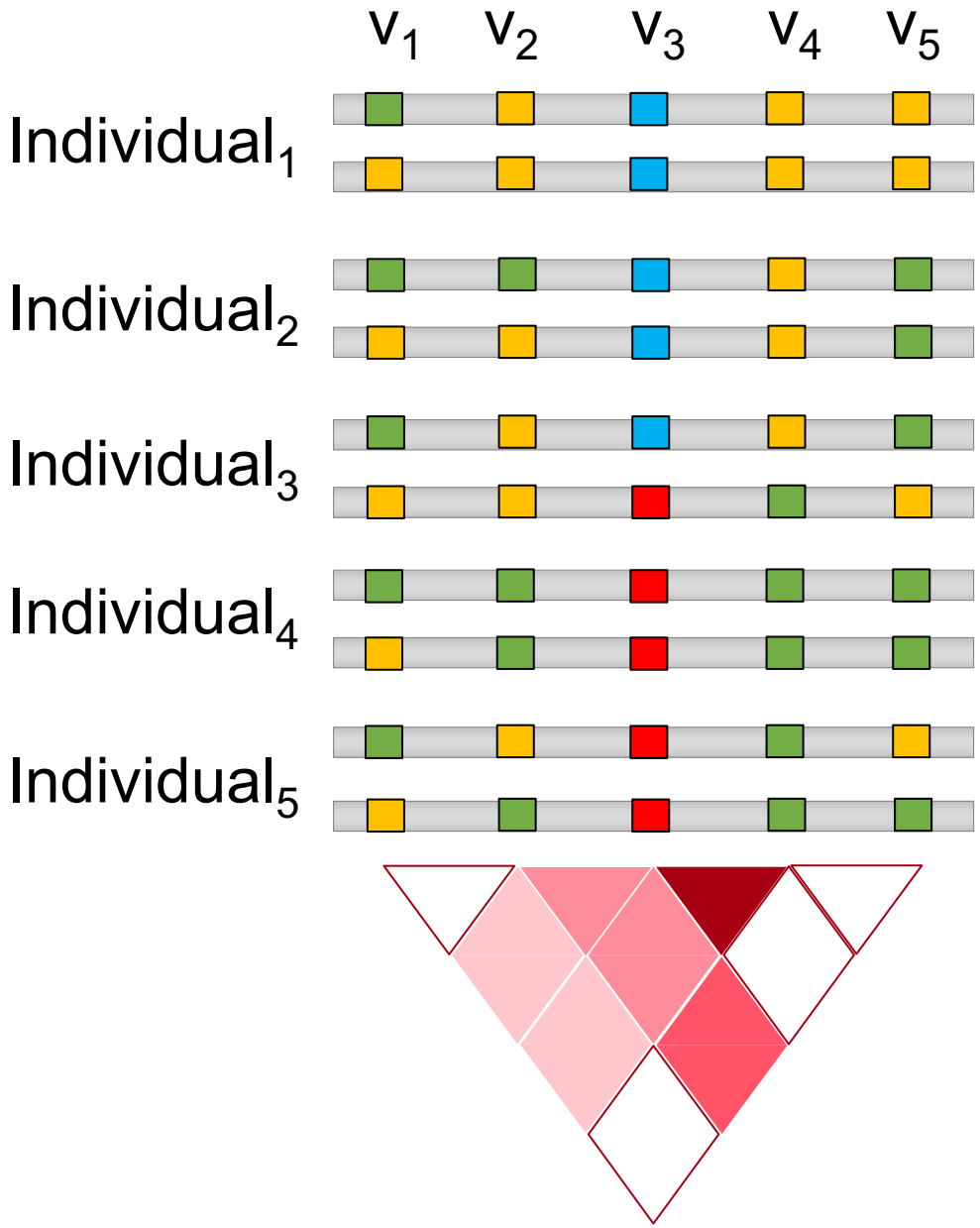
Linkage Disequilibrium

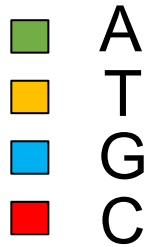


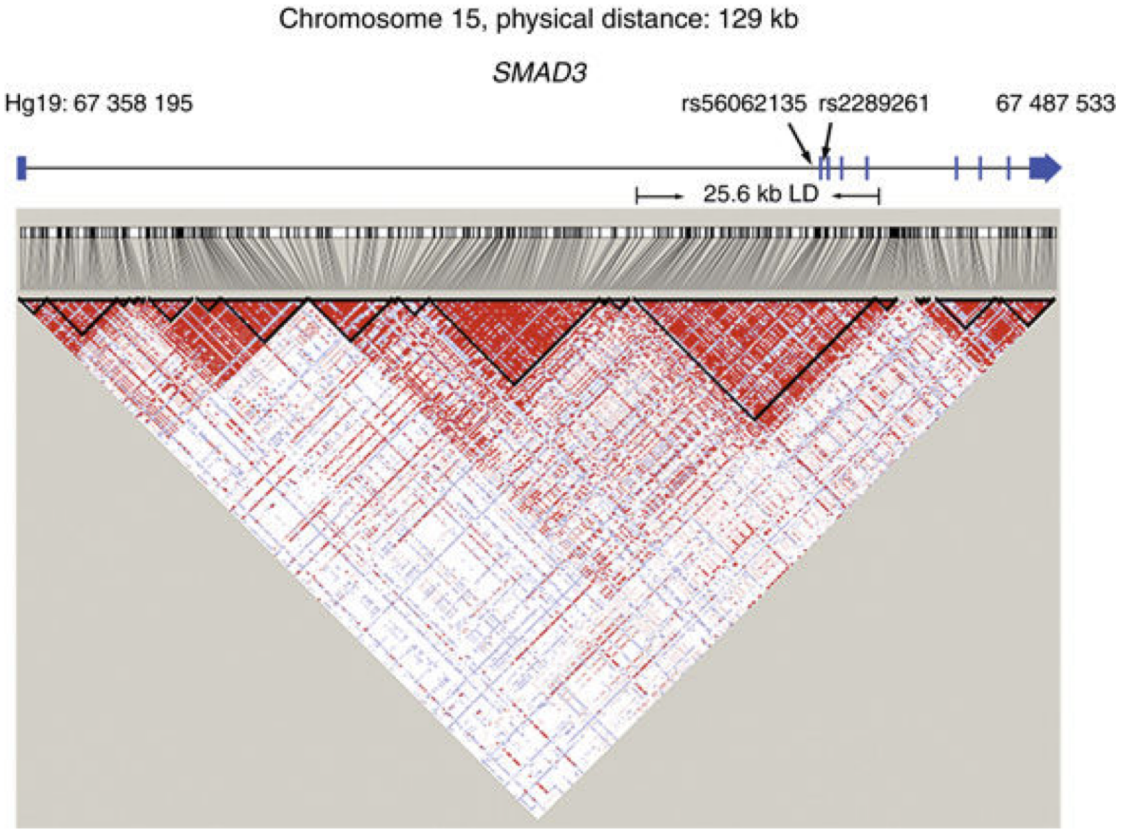
- Nonrandom association of alleles at different loci
- LD is a population-specific measurement

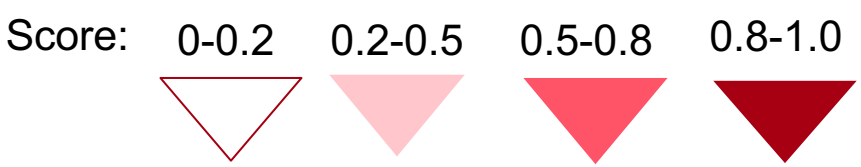




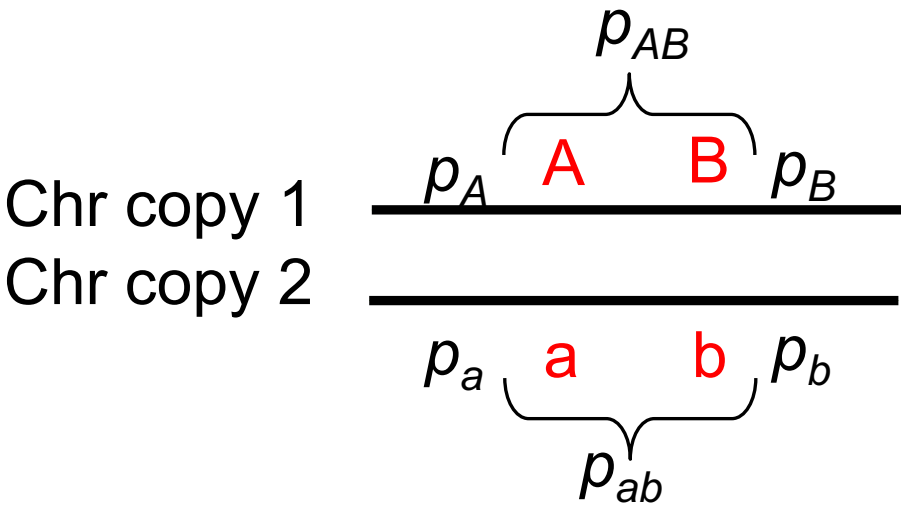

V1
V2
V3
V4
V5
Individual1
Individual2
Individual3
Individual4
Individual5


A
T
G
C


Chromosome 15, physical distance: 129 kb
SMAD3
Hg19: 67 358 195
rs56062135
rs2289261
67 487 533
25.6 kb LD


Score:
0-0.2
0.2-0.5
0.5-0.8
0.8-1.0

LD calculation



Allele	A	a	probability
B	p_{AB}	p_{aB}	p_B
b	p_{Ab}	p_{ab}	p_b
probability	p_A	p_a	1.0

Haplotype frequencies: p_{AB}, p_{ab}, p_{Ab}, p_{aB}
Allele frequencies: p_A, p_B, p_a, p_b
LD = observed - expected

$$D = p_{AB} - p_A p_B$$
$$|D'| = D/D_{max}$$
$$r^2 = D^2/p_A p_a p_B p_b$$

Interpreting LD values *D'* and r^2

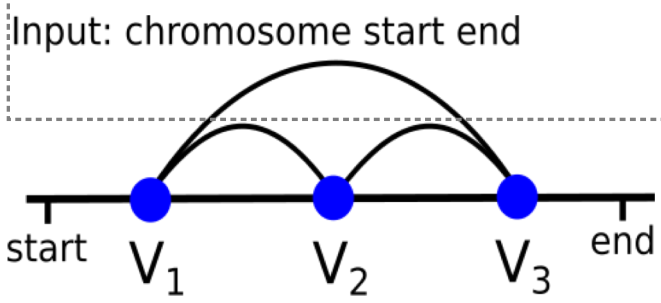
- High *D'* and high r^2
  - Tendency to the presence of only 2 haplotypes, small difference in allele frequency of coupled alleles
- High *D'* and low r^2
  - Tendency to the presence of only 3 haplotypes, different allele frequencies of the coupled alleles
- Low *D'* and low r^2
  - Tendency toward random coupling of alleles and presence of all 4 haplotypes

LDFeatureContainerAdaptor

```
# Create LDFeatureContainerAdaptor
my $ldfca = $registry->get_adaptor('human', 'variation', 'LDFeatureContainer');

# TODO fetch population object

# Compute LD in a region
# TODO fetch slice object for your region
my $ld_feature_container = $ldfca->fetch_by_Slice($slice, $population);
```



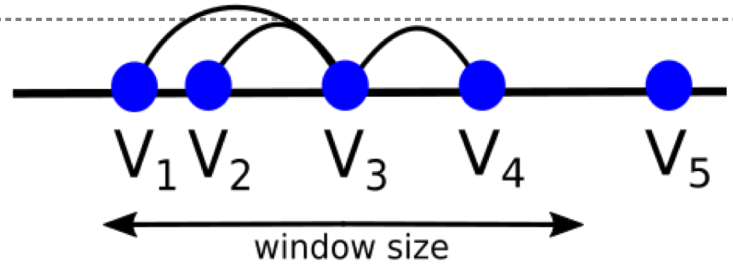
LDFeatureContainerAdaptor

```
# Create LDFeatureContainerAdaptor
my $ldfca = $registry->get_adaptor('human', 'variation', 'LDFeatureContainer');

# TODO fetch Population object

# Compute LD for a VariationFeature and all surrounding
# VariationFeatures that lie within a given window size
# max window size is XXX, default window size is XXX
# TODO fetch VariationFeature object
my $ld_feature_container = $ldfca->fetch_by_VariationFeature($vf, $population);
```

Input: Variant V_3, window size



LDFeatureContainer

- Container of pairwise LD values
- Calculated on the fly (using a script written in C)
- Use `$ldfc->get_all_ld_values` to get all computed LD values that are stored in the container $ldfc
- get_all_ld_values returns an array reference of hash references

Key	Value
variation1	VariationFeature object of one of the two variations used for the calculation
variation_name1	String value
variation2	VariationFeature object of the second variation used for the calculation
variation_name2	String value
r2	0 to 1
d_prime	0 to 1

```
foreach my $ld_result (@{$ldfc->get_all_ld_values}) {
  print $ld_result->{r2}, "\n";
…
}
```

LD populations

```
# Create Population adaptor
my $pa = $registry->get_adaptor('human', 'variation', 'Population');

my $ld_populations = $pa->fetch_all_LD_Populations();

foreach my $ld_population (@$ld_populations) {
    my $name = $ld_population->name;
    my $description = $ld_population->description;
    print "$name $description\n";
}
```

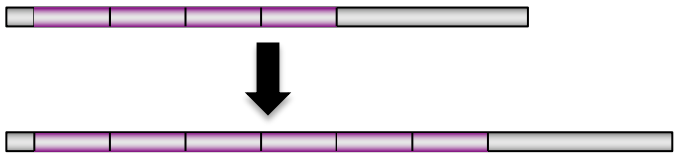
Exercise

- LD in a region `fetch_by_Slice($slice, $population)`
  - Compute all LD values for the region chromosome 12 between 11053716-11073715 and population 1000GENOMES:phase_3:CEU
  - Only print results where D' = 1 and r^2 = 1. Print D' (d_prime), r^2 (r2), variation_name1 and variation_name2
- LD for a variation feature `fetch_by_VariationFeature($vf, $population)`
  - Compute all LD values for rs7304579 and population 1000GENOMES:phase_3:YRI
  - Get the VariationFeature for rs7304579
  - Print all results. Print D' (d_prime), r^2 (r2), variation_name1 and variation_name2

Structural Variation

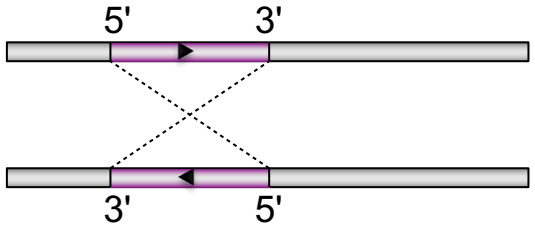
- Usually represents a large segment of variable DNA sequence (> 1kb)
- Different processes give rise to different types of structural variations:

Copy number variation (CNV)

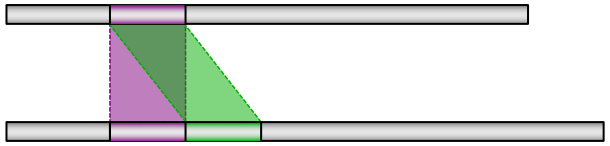


→ Allele type “Gain”

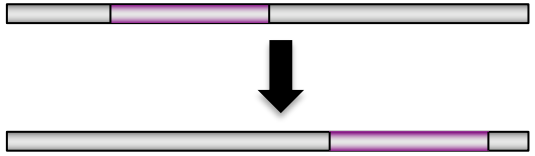
Inversion



Duplication



Translocation



Example of a Structural Variant

Structural variant: nsv917561

Variation class	CNV (SO:0001019)
Source	DGVa – Database of Genomic Variants Archive
Study	nstd75 – "International Standards for Cytogenomic Arrays Consortium (prenatal dataset)." PMID:21844811, PMID:20466091
Location	Chromosome 1:146987841-148359881 (forward strand)

Supporting evidence

Supporting evidence	Chr:bp (strand)	Allele type	Clinical significance	Individual name
nssv1605931	1:146987841-148359881 (+)	Gain	pathogenic	ISCA_ID_pn_1622
nssv1606072	1:146987841-148359881 (+)	Gain	-	ISCA_ID_pn_1746

Structural variant: nsv917561

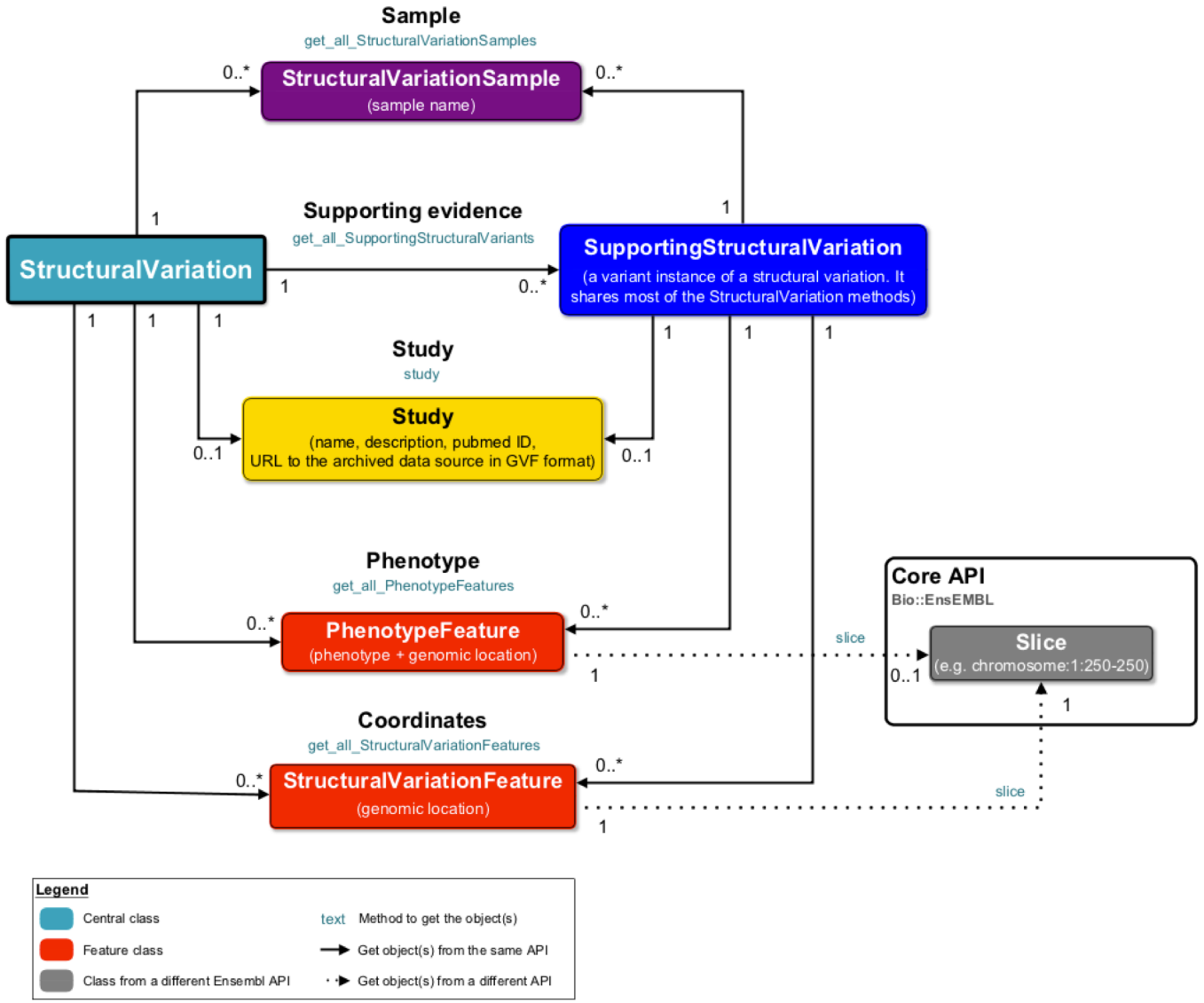
Variation class	CNV (SO:0001019)
Allele type(s)	gain (SO:0001742)
Source	DGVa - Database of Genomic Variants Archive
Study	nstd75 - International Standards for Cytogenomic Arrays Consortium (prenatal dataset) PMID:21844811 PMID:20466091
Alias	ISCA_VAR_pn_694
Clinical significance	
Location	Chromosome **1:146987841-148435812** (forward strand) | View in location tab
Genomic size	1,447,972 bp
About this structural variant	This structural variant overlaps 162 transcripts, is associated with 2 phenotypes and is supported by 2 pieces of evidence.

Explore this SV

http://www.ensembl.org/Homo_sapiens/StructuralVariation/Explore?sv=nsv917561

EnsEMBL Variation API Overview - StructuralVariation centered

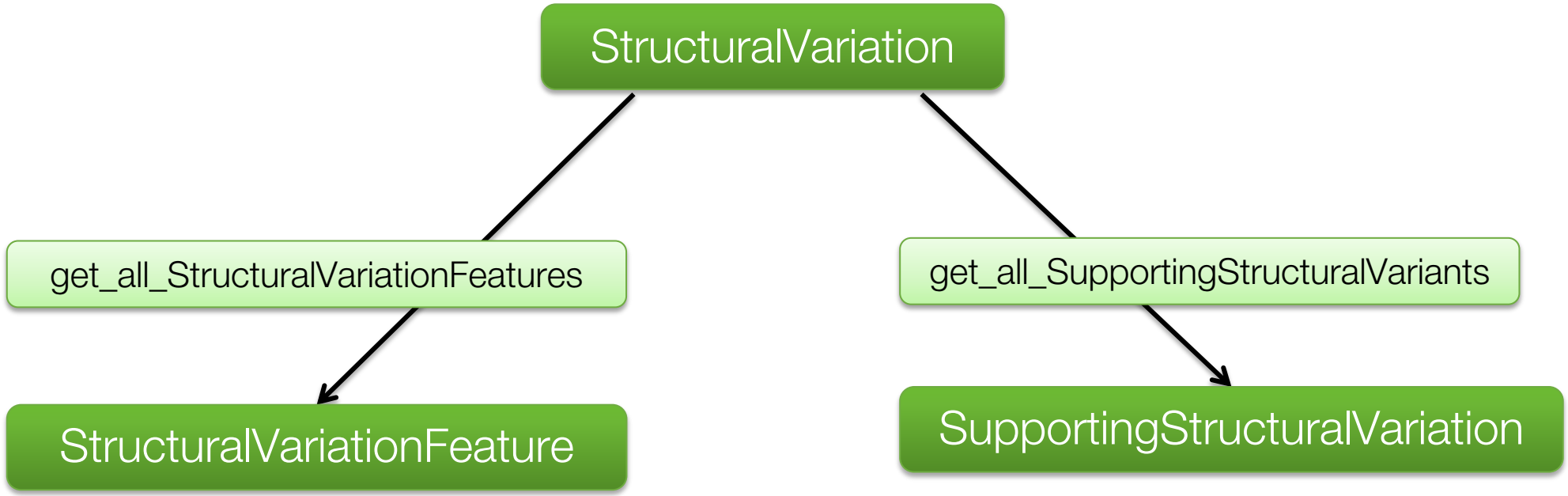
Bio::EnsEMBL::Variation



StructuralVariation

Method	Return type	Example value
`$sv->variation_name()`	String	esv214236
`$sv->var_class()`	String	CNV
`$sv->study()`	Study object	Bio::EnsEMBL::Variation::Study, undef if not defined
`$sv->get_all_SupportingStructuralVariants()`	Listref of objects	[Bio::EnsEMBL::Variation::SupportingStructuralVariation]

Supporting structural variation

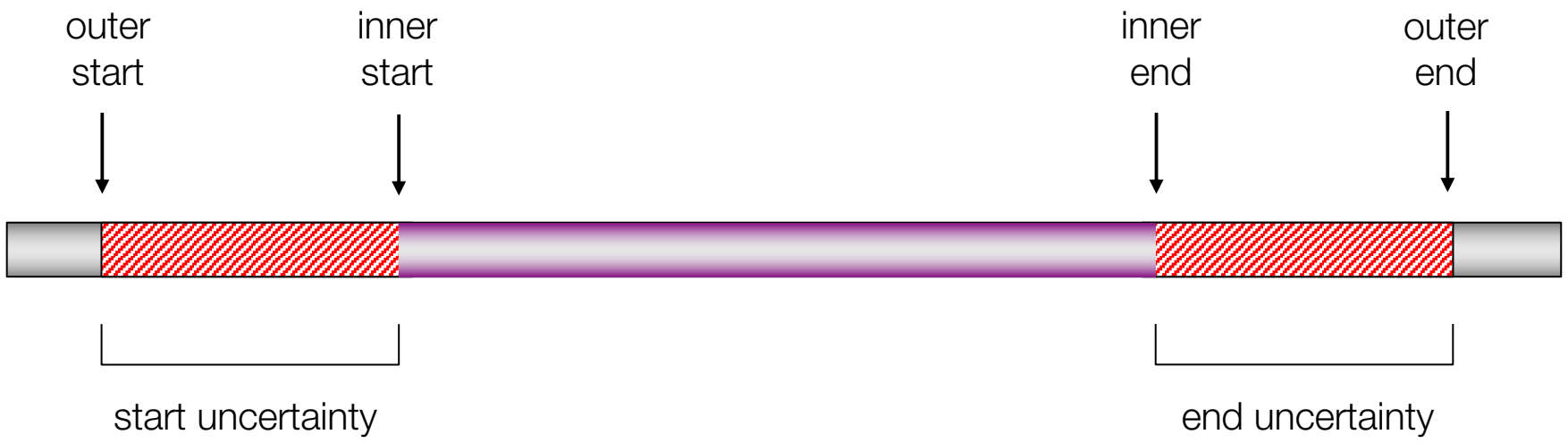


Region of the genome that a submitter has defined as containing structural variations.

Actual variant calls that were made within a study.

Structural variation coordinates

- Usually a structural variation has only a start and an end
- But sometimes, the breakpoint locations cannot be determined precisely



StructuralVariationFeature

- Can be retrieved from the StructuralVariationFeature adaptor and the StructuralVariation object

Attribute	Example value(s)	Method(s)
Structural variation name	esv214236	`$svf->variation_name()`
Chromosome	15, X	`$svf->seq_region_name()`

Coordinates		`$svf->start()`
		`$svf->end()`
		`$svf->seq_region_start()`
		`$svf->seq_region_end()` — "Classic" coordinates
		
		`$svf->outer_start()`
		`$svf->inner_start()`
		`$svf->inner_end()`
		`$svf->outer_end()` — SV specific coordinates

Class	CNV	`$svf->var_class()`

Exercise

- Fetch the following information for the structural variation esv275264 in human :
  - Structural variation class
  - Study name and description
  - Coordinates
- Fetch the names and the classes (sequence ontology term) of its supporting structural variations.

Ensembl Acknowledgements

The Entire Ensembl Team

Daniel R. Zerbino[1], Premanand Achuthan[1], Wasiu Akanni[1], M. Ridwan Amode[1], Daniel Barrell[1,2], Jyothish Bhai[1], Konstantinos Billis[1], Carla Cummins[1], Astrid Gall[1], Carlos García Giroń[1], Laurent Gil[1], Leo Gordon[1], Leanne Haggerty[1], Erin Haskell[1], Thibaut Hourlier[1], Osagie G. Izuogu[1], Sophie H. Janacek[1], Thomas Juettemann[1], Jimmy Kiang To[1], Matthew R. Laird[1], Ilias Lavidas[1], Zhicheng Liu[1], Jane E. Loveland[1], Thomas Maurel[1], William McLaren[1], Benjamin Moore[1], Jonathan Mudge[1], Daniel N. Murphy[1], Victoria Newman[1], Michael Nuhn[1], Denye Ogeh[1], Chuang Kee Ong[1], Anne Parker[1], Mateus Patricio[1], Harpreet Singh Riat[1], Helen Schuilenburg[1], Dan Sheppard[1], Helen Sparrow[1], Kieron Taylor[1], Anja Thormann[1], Alessandro Vullo[1], Brandon Walts[1], Amonida Zadissa[1], Adam Frankish[1], Sarah E. Hunt[1], Myrto Kostadima[1], Nicholas Langridge[1], Fergal J. Martin[1], Matthieu Muffato[1], Emily Perry[1], Magali Ruffier[1], Dan M. Staines[1], Stephen J. Trevanion[1], Bronwen L. Aken[1], Fiona Cunningham[1], Andrew Yates[1] and Paul Flicek[1,3]

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

Funding

wellcome

EMBL

National Human Genome Research Institute

SAVE THE TASMANIAN DEVIL PROGRAM APPEAL

Open Targets

BBSRC bioscience for the future

TRANSFORMING GENETIC MEDICINE INITIATIVE

MedBioinformatics

Co-funded by the European Union