

In [1]:

```
import pandas as pd
import pandas_profiling
import numpy as np
```

In [5]:

```
df=pd.read_csv("raw_data/merged_withVEP.csv")
pandas_profiling.ProfileReport(df)
```

Out[5]:

Overview

Dataset info

Number of variables	112
Number of observations	15433
Total Missing (%)	23.1%
Total size in memory	13.3 MiB
Average record size in memory	904.0 B

Variables types

Numeric	0
Categorical	109
Date	0
Text (Unique)	1
Rejected	2

Warnings

- AFR_Allele_frequency(1000_Genomes) has a high cardinality: 236 distinct values

Warning
- AMR_Allele_frequency(1000_Genomes) has a high cardinality: 155 distinct values

Warning
- Abbrev_AA_change(ENIGMA) has 845 / 5.5% missing values

Missing
- Abbrev_AA_change(ENIGMA) has a high cardinality: 187 distinct values

Warning
- Allele_frequency(1000_Genomes) has a high cardinality: 337 distinct values

Warning
- Allele_frequency(ExAC) has a high cardinality: 425 distinct values

Warning

- BIC_Designation(BIC) has a high cardinality: 2960 distinct values
Warning
- BIC_Nomenclature(ENIGMA) has a high cardinality: 1030 distinct values
Warning
- BIC_identifier(exLOVD) has a high cardinality: 217 distinct values
Warning
- ClinVarAccession(ENIGMA) has a high cardinality: 1031 distinct values
Warning
- Clinical_significance_citations(ENIGMA) has 822 / 5.3% missing values
Missing
- Co_occurrence_LR(exLOVD) has a high cardinality: 140 distinct values
Warning
- Comment_on_clinical_significance(ENIGMA) has a high cardinality: 529 distinct values
Warning
- Date_Last_Updated(ClinVar) has a high cardinality: 3475 distinct values
Warning
- EAS_Allele_frequency(1000_Genomes) has a high cardinality: 142 distinct values
Warning
- EUR_Allele_frequency(1000_Genomes) has a high cardinality: 155 distinct values
Warning
- Ethnicity(BIC) has a high cardinality: 1017 distinct values
Warning
- Functional_analysis_result(LOVD) has a high cardinality: 199 distinct values
Warning
- Functional_analysis_technique(LOVD) has a high cardinality: 276 distinct values
Warning
- HGVS(ClinVar) has a high cardinality: 6886 distinct values
Warning
- HGVS_cDNA(ENIGMA) has a high cardinality: 1031 distinct values
Warning
- HGVS_protein(ENIGMA) has a high cardinality: 201 distinct values
Warning
- Literature_citation(BIC) has a high cardinality: 195 distinct values
Warning
- Minor_allele_frequency(ESP) has a high cardinality: 205 distinct values
Warning
- Number_of_family_member_carrying_mutation(BIC) has a high cardinality: 69 distinct values
Warning
- Patient_nationality(BIC) has a high cardinality: 423 distinct values
Warning
- Posterior_probability(exLOVD) has a high cardinality: 89 distinct values
Warning
- SAS_Allele_frequency(1000_Genomes) has a high cardinality: 149 distinct values
Warning
- SCV(ClinVar) has a high cardinality: 6886 distinct values
Warning
- Segregation_LR(exLOVD) has a high cardinality: 107 distinct values
Warning
- Source has a high cardinality: 81 distinct values
Warning

- Source has a high cardinality: 81 distinct values **Warning**
- Submitter(ClinVar) has a high cardinality: 358 distinct values **Warning**
- Sum_family_LR(exLOVD) has a high cardinality: 144 distinct values **Warning**
- URL(ENIGMA) has 822 / 5.3% missing values **Missing**
- URL(ENIGMA) has a high cardinality: 210 distinct values **Warning**
- VEP_AA_MAF has 14184 / 91.9% missing values **Missing**
- VEP_AA_MAF has a high cardinality: 137 distinct values **Warning**
- VEP_AFR_MAF has 10938 / 70.9% missing values **Missing**
- VEP_AFR_MAF has a high cardinality: 430 distinct values **Warning**
- VEP_AMR_MAF has 10938 / 70.9% missing values **Missing**
- VEP_AMR_MAF has a high cardinality: 289 distinct values **Warning**
- VEP_Allele has a high cardinality: 192 distinct values **Warning**
- VEP_Amino_acids has 5851 / 37.9% missing values **Missing**
- VEP_Amino_acids has a high cardinality: 4212 distinct values **Warning**
- VEP_BIOTYPE has a high cardinality: 83 distinct values **Warning**
- VEP_CDS_position has 5743 / 37.2% missing values **Missing**
- VEP_CDS_position has a high cardinality: 6604 distinct values **Warning**
- VEP_CLIN_SIG has 8560 / 55.5% missing values **Missing**
- VEP_CLIN_SIG has a high cardinality: 77 distinct values **Warning**
- VEP_Codons has 5848 / 37.9% missing values **Missing**
- VEP_Codons has a high cardinality: 5948 distinct values **Warning**
- VEP_Consequence has a high cardinality: 991 distinct values **Warning**
- VEP_DISTANCE has 3270 / 21.2% missing values **Missing**
- VEP_DISTANCE has a high cardinality: 7994 distinct values **Warning**
- VEP_EAS_MAF has 10967 / 71.1% missing values **Missing**
- VEP_EAS_MAF has a high cardinality: 225 distinct values **Warning**
- VEP_EA_MAF has 14184 / 91.9% missing values **Missing**
- VEP_EA_MAF has a high cardinality: 92 distinct values **Warning**
- VEP_EUR_MAF has 10939 / 70.9% missing values **Missing**
- VEP_EUR_MAF has a high cardinality: 264 distinct values **Warning**
- VEP_EXON has 5369 / 34.8% missing values **Missing**
- VEP_EXON has a high cardinality: 135 distinct values **Warning**
- VEP_ExAC_AFR_MAF has 10859 / 70.4% missing values **Missing**
- VEP_ExAC_AFR_MAF has a high cardinality: 574 distinct values **Warning**
- VEP_ExAC_AMR_MAF has 10862 / 70.4% missing values **Missing**
- VEP_ExAC_AMR_MAF has a high cardinality: 449 distinct values **Warning**
- VEP_ExAC_Adj_MAF has 10856 / 70.3% missing values **Missing**
- VEP_ExAC_Adj_MAF has a high cardinality: 1774 distinct values **Warning**
- VEP_ExAC_EAS_MAF has 10857 / 70.3% missing values **Missing**
- VEP_ExAC_EAS_MAF has a high cardinality: 275 distinct values **Warning**
- VEP_ExAC_FIN_MAF has 10906 / 70.7% missing values **Missing**
- VEP_ExAC_FIN_MAF has a high cardinality: 161 distinct values **Warning**
- VEP_ExAC_MAF has 10856 / 70.3% missing values **Missing**
- VEP_ExAC_MAF has a high cardinality: 600 distinct values **Warning**

- VEP_ExAC_MAF has a high cardinality: 690 distinct values **Warning**
- VEP_ExAC_NFE_MAF has 10857 / 70.3% missing values **Missing**
- VEP_ExAC_NFE_MAF has a high cardinality: 885 distinct values **Warning**
- VEP_ExAC_OTH_MAF has 10861 / 70.4% missing values **Missing**
- VEP_ExAC_OTH_MAF has a high cardinality: 149 distinct values **Warning**
- VEP_ExAC_SAS_MAF has 10857 / 70.3% missing values **Missing**
- VEP_ExAC_SAS_MAF has a high cardinality: 509 distinct values **Warning**
- VEP_Existing_variation has 2944 / 19.1% missing values **Missing**
- VEP_Existing_variation has a high cardinality: 8619 distinct values **Warning**
- VEP_Feature has a high cardinality: 89 distinct values **Warning**
- VEP_GMAF has 10912 / 70.7% missing values **Missing**
- VEP_GMAF has a high cardinality: 582 distinct values **Warning**
- VEP_Gene has a high cardinality: 80 distinct values **Warning**
- VEP_HGNC_ID has a high cardinality: 68 distinct values **Warning**
- VEP_HGVSc has 15433 / 100.0% missing values **Missing**
- VEP_HGVSc has constant value **Rejected**
- VEP_HGVSp has 15433 / 100.0% missing values **Missing**
- VEP_HGVSp has constant value **Rejected**
- VEP_HIGH_INF_POS has 15425 / 99.9% missing values **Missing**
- VEP_IMPACT has a high cardinality: 351 distinct values **Warning**
- VEP_INTRON has 6045 / 39.2% missing values **Missing**
- VEP_INTRON has a high cardinality: 160 distinct values **Warning**
- VEP_MOTIF_NAME has 15425 / 99.9% missing values **Missing**
- VEP_MOTIF_POS has 15425 / 99.9% missing values **Missing**
- VEP_MOTIF_SCORE_CHANGE has 15425 / 99.9% missing values **Missing**
- VEP_PHENO has 8110 / 52.5% missing values **Missing**
- VEP_PUBMED has 15127 / 98.0% missing values **Missing**
- VEP_PUBMED has a high cardinality: 110 distinct values **Warning**
- VEP_PolyPhen has 10317 / 66.9% missing values **Missing**
- VEP_PolyPhen has a high cardinality: 2438 distinct values **Warning**
- VEP_Protein_position has 5743 / 37.2% missing values **Missing**
- VEP_Protein_position has a high cardinality: 4554 distinct values **Warning**
- VEP_REFSEQ_MATCH has a high cardinality: 51 distinct values **Warning**
- VEP_SAS_MAF has 10967 / 71.1% missing values **Missing**
- VEP_SAS_MAF has a high cardinality: 253 distinct values **Warning**
- VEP_SIFT has 10320 / 66.9% missing values **Missing**
- VEP_SIFT has a high cardinality: 1519 distinct values **Warning**
- VEP_SOMATIC has 14604 / 94.6% missing values **Missing**
- VEP_SYMBOL has a high cardinality: 68 distinct values **Warning**
- VEP_SYMBOL_SOURCE has a high cardinality: 66 distinct values **Warning**
- VEP_TSL has a high cardinality: 75 distinct values **Warning**
- VEP_cDNA_position has 5369 / 34.8% missing values **Missing**
- VEP_cDNA_position has a high cardinality: 6899 distinct values **Warning**

Warning

- Variant_frequency(LOVD) has a high cardinality: 80 distinct values
- polyPhen2_result(ESP) has a high cardinality: 211 distinct values

Warning

Warning

Variables

AFR_Allele_frequency(1000_Genomes)

Categorical

Distinct count	236
Unique (%)	1.5%
Missing (%)	0.0%
Missing (n)	0
	-12079
	['0.0']1696
	['0.0008']574
Other values (233)	1084

Toggle details

AMR_Allele_frequency(1000_Genomes)

Categorical

Distinct count	155
Unique (%)	1.0%
Missing (%)	0.0%
Missing (n)	0
	-12079
	['0.0']2308
	['0.0014']423
Other values (152)	623

Toggle details

Abbrev_AA_change(ENIGMA)

Categorical

Distinct count	187			
Unique (%)	1.3%			
Missing (%)	5.5%			
Missing (n)	845			
	-			14403
	L1764P	1		
	I1766S	1		
	Other values (183)	183		
	(Missing)	845		
Toggle details				

Allele_Origin(ClinVar)				
Categorical				
Distinct count	10			
Unique (%)	0.1%			
Missing (%)	0.0%			
Missing (n)	0			
	-			8548
	['germline']			6570
	['germline unknown']	272		
	Other values (7)	43		
Toggle details				

Allele_frequency(1000_Genomes)				
Categorical				
Distinct count	337			
Unique (%)	2.2%			
Missing (%)	0.0%			
Missing (n)	0			
	-			12079
	[0.000199681]	1512		
	[0.000399361]	378		
	Other values (334)	1464		
Toggle details				

--	--	--	--	--

Allele_frequency(ExAC)

Categorical

Distinct count425

Unique (%)2.8%

Missing (%)0.0%

Missing (n)0

-	12428
[8.237e-06]	601
[8.236e-06]	599
Other values (422)	1805

Toggle details

Allele_origin(ENIGMA)

Categorical

Distinct count2

Unique (%)0.0%

Missing (%)0.0%

Missing (n)0

-	14403
Germline	1030

Toggle details

Assertion_method(ENIGMA)

Categorical

Distinct count2

Unique (%)0.0%

Missing (%)0.0%

Missing (n)0

-	14403
ENIGMA BRCA1/2 Classification Crite...	1030

Toggle details

Assertion_method_citation(ENIGMA)

Categorical

Distinct count	2		
Unique (%)	0.0%		
Missing (%)	0.0%		
Missing (n)	0		
	-		14403
http://enigmaconsortium.org/docume...	1030		
Toggle details			

BIC_Designation(BIC)			
Categorical			
Distinct count	2960		
Unique (%)	19.2%		
Missing (%)	0.0%		
Missing (n)	0		
	-		12433
	['Y1655X']	3	
	['IVS17+1G>A']	3	
	Other values (2957)	2994	
Toggle details			

BIC_Nomenclature(ENIGMA)			
Categorical			
Distinct count	1030		
Unique (%)	6.7%		
Missing (%)	0.0%		
Missing (n)	0		
	-		14403
	IVS3+1936C>G	2	
	IVS19+2230C>G	1	
	Other values (1027)	1027	
Toggle details			

BIC_identifier(exLOVD)			
Categorical			
Distinct count	217		

Unique (%)	1.4%			
Missing (%)	0.0%			
Missing (n)	0			
		-		15217
	['6566A>G']	1		
	['5104T>C']	1		
	Other values (214)	214		
Toggle details				

<div> <div>ClinVarAccession(ENIGMA)</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>1031</div> </div> <div> <div>Unique (%)</div> <div>6.7%</div> </div> <div> <div>Missing (%)</div> <div>0.0%</div> </div> <div> <div>Missing (n)</div> <div>0</div> </div> </div>				
		-		14403
	SCV000245069	1		
	SCV000244895	1		
	Other values (1028)	1028		
Toggle details				

<div> <div>Clinical_Significance(ClinVar)</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>41</div> </div> <div> <div>Unique (%)</div> <div>0.3%</div> </div> <div> <div>Missing (%)</div> <div>0.0%</div> </div> <div> <div>Missing (n)</div> <div>0</div> </div> </div>				
		-		8548
	['Uncertain_significance']	2602		
	['Pathogenic']	1932		
	Other values (38)	2351		
Toggle details				

<div> <div>Clinical_classification(BIC)</div> <div>Categorical</div> </div>				
---	--	--	--	--

Distinct count 8

Unique (%) 0.1%

Missing (%) 0.0%

Missing (n) 0

-	12433
['Pending']	1637
['Class 5']	1290
Other values (5)	73

Toggle details

Clinical_importance(BIC)

Categorical

Distinct count 7

Unique (%) 0.0%

Missing (%) 0.0%

Missing (n) 0

-	12433
['yes']	1459
['unknown']	1417
Other values (4)	124

Toggle details

Clinical_significance(ENIGMA)

Categorical

Distinct count 3

Unique (%) 0.0%

Missing (%) 0.0%

Missing (n) 0

-	14403
Benign	986
Pathogenic	44

Toggle details

Clinical_significance_citations(ENIGMA)

Categorical

Distinct count	8		
Unique (%)	0.1%		
Missing (%)	5.3%		
Missing (n)	822		
	-		14403
	PMID:21990134	189	
	PMID:23108138	9	
	Other values (4)	10	
	(Missing)	822	

Toggle details

Co_occurrence_LR(exLOVD)

Categorical

Distinct count	140		
Unique (%)	0.9%		
Missing (%)	0.0%		
Missing (n)	0		
	-		15217
	['-']	30	
	['1.17']	5	
	Other values (137)	181	

Toggle details

Collection_method(ENIGMA)

Categorical

Distinct count	2		
Unique (%)	0.0%		
Missing (%)	0.0%		
Missing (n)	0		
	-		14403
	Curation	1030	

Toggle details

Combined_prior_probablility(exLOVD)

Categorical

Distinct count	13			
Unique (%)	0.1%			
Missing (%)	0.0%			
Missing (n)	0			
	-			15217
	['0.02']	119		
	['0.81']	23		
	Other values (10)	74		
Toggle details				

Comment_on_clinical_significance(ENIGMA)				
Categorical				
Distinct count	529			
Unique (%)	3.4%			
Missing (%)	0.0%			
Missing (n)	0			
	-			14403
	IARC class based on posterior proba...	36		
	Class 1 not pathogenic based on freq...	28		
	Other values (526)	966		
Toggle details				

Condition_ID_type(ENIGMA)				
Categorical				
Distinct count	2			
Unique (%)	0.0%			
Missing (%)	0.0%			
Missing (n)	0			
	-			14403
	OMIM	1030		
Toggle details				

Condition_ID_value(ENIGMA)				
Categorical				
Distinct count	3			

Unique (%)	0.0%				
Missing (%)	0.0%				
Missing (n)	0				
		-		14403	
BREAST-OVARIAN CANCER FAMILI...	533				
BREAST-OVARIAN CANCER FAMILI...	497				
Toggle details					

Condition_category(ENIGMA) Categorical					
Distinct count	2				
Unique (%)	0.0%				
Missing (%)	0.0%				
Missing (n)	0				
		-		14403	
	Disease	1030			
Toggle details					

Date_Last_Updated(ClinVar) Categorical					
Distinct count	3475				
Unique (%)	22.5%				
Missing (%)	0.0%				
Missing (n)	0				
		-		8548	
	['2015-01-12']	757			
	['2002-05-29']	286			
	Other values (3472)		5842		
Toggle details					

Date_last_evaluated(ENIGMA) Categorical					
Distinct count	3				
Unique (%)	0.0%				
Missing (%)	0.0%				

Missing (n)	0				
		-		14403	
		1/12/15	822		
		8/10/15	208		
Toggle details					

EAS_Allele_frequency(1000_Genomes)					
Categorical					
Distinct count	142				
Unique (%)	0.9%				
Missing (%)	0.0%				
Missing (n)	0				
		-		12079	
		['0.0']	2386		
		['0.001']	412		
		Other values (139)	556		
Toggle details					

EUR_Allele_frequency(1000_Genomes)					
Categorical					
Distinct count	155				
Unique (%)	1.0%				
Missing (%)	0.0%				
Missing (n)	0				
		-		12079	
		['0.0']	2453		
		['0.001']	326		
		Other values (152)	575		
Toggle details					

Ethnicity(BIC)					
Categorical					
Distinct count	1017				
Unique (%)	6.6%				
Missing (%)	0.0%				

Missing (%)	0.0%			
Missing (n)	0			
		-		12433
		['-']	583	
		['Western European']	213	
		Other values (1014)	2204	
Toggle details				

Functional_analysis_result(LOVD)				
Categorical				
Distinct count	199			
Unique (%)	1.3%			
Missing (%)	0.0%			
Missing (n)	0			
		-		12707
		['-']	1908	
		['predicted neutral -']	115	
		Other values (196)	703	
Toggle details				

Functional_analysis_technique(LOVD)				
Categorical				
Distinct count	276			
Unique (%)	1.8%			
Missing (%)	0.0%			
Missing (n)	0			
		-		12707
		['-']	1909	
		['Protein likelihood ratio model']	88	
		Other values (273)	729	
Toggle details				

Gene_symbol(ENIGMA)				
Categorical				
Distinct count	3			

Unique (%)	0.0%				
Missing (%)	0.0%				
Missing (n)	0				
		-		14403	
		BRCA1	533		
		BRCA2	497		
Toggle details					

Genomic_Coordinate(ENIGMA)
Categorical, Unique
First 3 values
chr17:43051120:G>A
chr17:43122039:G>C
chr13:32327584:C>A
Last 3 values
chr17:43097519:G>T
chr17:43071092:ATCCATCTCCCTGTGA>C
chr17:43092777:C>A
Toggle details

Germline_or_Somatic(BIC)

Categorical

Distinct count

27

Unique (%)

0.2%

Missing (%)

0.0%

Missing (n)

0

-

12433

['G']

2720

['-|G']

86

Other values (24)

194

Toggle details

HGVS(ClinVar)	
Categorical	
Distinct count	6886
Unique (%)	44.6%
Missing (%)	0.0%

Missing (n)	0		
	-		8548
['NM_000059.3.c.6059_6062delAACA']	1		
['NM_007294.3.c.3973delA']	1		
Other values (6883)			6883
Toggle details			

<div> <div>HGVS_cDNA(ENIGMA)</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>1031</div> </div> <div> <div>Unique (%)</div> <div>6.7%</div> </div> <div> <div>Missing (%)</div> <div>0.0%</div> </div> <div> <div>Missing (n)</div> <div>0</div> </div> </div>			
	-		14403
c.5054C>T	1		
c.316+361A>G	1		
Other values (1028)		1028	
Toggle details			

<div> <div>HGVS_protein(ENIGMA)</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>201</div> </div> <div> <div>Unique (%)</div> <div>1.3%</div> </div> <div> <div>Missing (%)</div> <div>0.0%</div> </div> <div> <div>Missing (n)</div> <div>0</div> </div> </div>			
	-		14403
p.?	811		
p.(=)	21		
Other values (198)		198	
Toggle details			

<div> <div>IARC_class(exLOVD)</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>6</div> </div> <div> <div>Unique (%)</div> <div>0.0%</div> </div> </div>			
--	--	--	--

Unique (%)	0.0%		
Missing (%)	0.0%		
Missing (n)	0		
	-		15217
['1 - Not pathogenic or of no clinical s...			
		135	
['5 - Definitely pathogenic']			
		38	
Other values (3)			
		43	
			Toggle details

Literature_citation(BIC)			
Categorical			
Distinct count	195		
Unique (%)	1.3%		
Missing (%)	0.0%		
Missing (n)	0		
	-		12433
['-']			
		2567	
['Unpublished -']			
		80	
Other values (192)			
		353	
			Toggle details

Literature_source(exLOVD)			
Categorical			
Distinct count	17		
Unique (%)	0.1%		
Missing (%)	0.0%		
Missing (n)	0		
	-		15217
['Easton DF et al. Am J Hum Genet 8...			
		121	
['Spearman AD et al. J Clin Oncol 26...			
		17	
Other values (14)			
		78	
			Toggle details

Method(ClinVar)			
Categorical			

Distinct count 13

Unique (%) 0.1%

Missing (%) 0.0%

Missing (n) 0

-	8548
['clinical_testing']	5659
['curation']	762
Other values (10)	464

[Toggle details](#)

Minor_allele_frequency(ESP)

Categorical

Distinct count 205

Unique (%) 1.3%

Missing (%) 0.0%

Missing (n) 0

-	14746
['0.0116 0.0 0.0077']	195
['0.0 0.0227 0.0077']	121
Other values (202)	371

[Toggle details](#)

Missense_analysis_prior_probability(exLOVD)

Categorical

Distinct count 8

Unique (%) 0.1%

Missing (%) 0.0%

Missing (n) 0

-	15217
['0.02']	126
['-']	27
Other values (5)	63

[Toggle details](#)

Mutation_type(BIC)

Categorical

Categorical

Distinct count 21
Unique (%) 0.1%
Missing (%) 0.0%
Missing (n) 0

-	12433
['M']	1106
['F']	1048
Other values (18)	846

Toggle details

Number_of_family_member_carrying_mutation(BIC)

Categorical

Distinct count 69
Unique (%) 0.4%
Missing (%) 0.0%
Missing (n) 0

-	12433
['-']	2631
['1']	80
Other values (66)	289

Toggle details

Origin_of_variant(LOVD)

Categorical

Distinct count 8
Unique (%) 0.1%
Missing (%) 0.0%
Missing (n) 0

-	12707
['Unknown']	1711
['In vitro (cloned)']	416
Other values (5)	599

Toggle details

Patient_nationality(BIC)

Categorical

Distinct count	423
Unique (%)	2.7%
Missing (%)	0.0%
Missing (n)	0

-	12433
['-']	1756
['Not Specified -']	59
Other values (420)	1185

Toggle details

Posterior_probability(exLOVD)

Categorical

Distinct count	89
Unique (%)	0.6%
Missing (%)	0.0%
Missing (n)	0

-	15217
['0.00000']	64
['0.000']	10
Other values (86)	142

Toggle details

Reference_sequence(ENIGMA)

Categorical

Distinct count	3
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

-	14403
NM_007294.3	533
NM_000059.3	497

Toggle details

SAS_Allele_frequency(1000_Genomes)

Categorical

Distinct count	149
Unique (%)	1.0%
Missing (%)	0.0%
Missing (n)	0

-	12079
['0.0']	2342
['0.001']	403
Other values (146)	609

[Toggle details](#)

SCV(ClinVar)

Categorical

Distinct count	6886
Unique (%)	44.6%
Missing (%)	0.0%
Missing (n)	0

-	8548
['SCV000187006 SCV000144456']	1
['SCV000210228']	1
Other values (6883)	6883

[Toggle details](#)

Segregation_LR(exLOVD)

Categorical

Distinct count	107
Unique (%)	0.7%
Missing (%)	0.0%
Missing (n)	0

-	15217
['1']	60
['-']	42
Other values (104)	114

[Toggle details](#)

Source

Categorical

Distinct count 81

Unique (%) 0.5%

Missing (%) 0.0%

Missing (n) 0

ClinVar	4881
1000_Genomes	2718
ExAC	1660
Other values (78)	6174

[Toggle details](#)

Submitter(ClinVar)

Categorical

Distinct count 358

Unique (%) 2.3%

Missing (%) 0.0%

Missing (n) 0

-	8548
['Ambry_Genetics']	1077
['Evidence-based_Network_for_the_In...]	762
Other values (355)	5046

[Toggle details](#)

Sum_family_LR(exLOVD)

Categorical

Distinct count 144

Unique (%) 0.9%

Missing (%) 0.0%

Missing (n) 0

-	15217
['-']	55
['2.89']	3
Other values (141)	158

URL(ENIGMA)

Categorical

Distinct count	210		
Unique (%)	1.4%		
Missing (%)	5.3%		
Missing (n)	822		
	-		14403
http://hci-exlovd.hci.utah.edu/variant...	1		
http://hci-exlovd.hci.utah.edu/variant...	1		
Other values (206)	206		
(Missing)	822		

Toggle details

VEP_AA_MAF

Categorical

Distinct count	137		
Unique (%)	11.0%		
Missing (%)	91.9%		
Missing (n)	14184		
	G:0	161	
	T:0	157	
	C:0	118	
Other values (133)	813		
(Missing)			14184

Toggle details

VEP_AFR_MAF

Categorical

Distinct count	430		
Unique (%)	9.6%		
Missing (%)	70.9%		
Missing (n)	10938		
	A:0-0000	579	
	T:0-0000	575	
	C:0-0000	507	

G:0-0000		507	
Other values (426)		2834	
(Missing)			10938
			Toggle details

VEP_AMR_MAF

Categorical

Distinct count	289
Unique (%)	6.4%
Missing (%)	70.9%
Missing (n)	10938

A:0-0000	776	
T:0-0000	774	
G:0-0000	729	
Other values (285)	2216	
(Missing)		10938

[Toggle details](#)

VEP_APPRIS

Categorical

Distinct count	43
Unique (%)	0.3%
Missing (%)	0.0%
Missing (n)	0

[" "7/9"]["P1 "2/9"]	4469	
[" "35/39"]["alternative2 "3/39"]["P...	1884	
[" "8/10"]["P1 "2/10"]	1851	
Other values (40)		7229

[Toggle details](#)

VEP_Allele

Categorical

Distinct count	192
Unique (%)	1.2%
Missing (%)	0.0%

Missing (n)	0		
		G	3687
		T	3671
		A	3617
	Other values (189)		4458
Toggle details			

<div>VEP_Amino_acids</div> <div>Categorical</div> <div> <div> <div>Distinct count</div> <div>4212</div> </div> <div> <div>Unique (%)</div> <div>44.0%</div> </div> <div> <div>Missing (%)</div> <div>37.9%</div> </div> <div> <div>Missing (n)</div> <div>5851</div> </div> </div> <div> <div> <div>["S" "7/9"]</div> <div>[" " "2/9"]</div> <div>42</div> </div> <div> <div>["K/N" "7/9"]</div> <div>[" " "2/9"]</div> <div>37</div> </div> <div> <div>["I/V" "7/9"]</div> <div>[" " "2/9"]</div> <div>37</div> </div> <div> <div>Other values (4208)</div> <div></div> <div>9466</div> </div> <div> <div>(Missing)</div> <div>5851</div> <div></div> </div> </div> <div>Toggle details</div>			
--	--	--	--

<div>VEP_BIOTYPE</div> <div>Categorical</div> <div> <div> <div>Distinct count</div> <div>83</div> </div> <div> <div>Unique (%)</div> <div>0.5%</div> </div> <div> <div>Missing (%)</div> <div>0.0%</div> </div> <div> <div>Missing (n)</div> <div>0</div> </div> </div> <div> <div> <div>["protein_coding" "8/9"]</div> <div>["processed_...</div> <div>2863</div> </div> <div> <div>["protein_coding" "34/39"]</div> <div>["nonsens...</div> <div>1518</div> </div> <div> <div>["protein_coding" "8/10"]</div> <div>["processed...</div> <div>1033</div> </div> <div> <div>Other values (80)</div> <div></div> <div>10019</div> </div> </div> <div>Toggle details</div>			
--	--	--	--

<div>VEP_CDS_position</div> <div>Categorical</div> <div> <div> <div>Distinct count</div> <div>6604</div> </div> </div>			
--	--	--	--

Unique (%)	68.2%		
Missing (%)	37.2%		
Missing (n)	5743		
[[{"1501" 7/28}][{"5/28"}][{"286" 4/2...]]	7		
[[{"2957" 7/10}][{"3/10"}]]	7		
[[{"25/39"}][{"1893" 8/39}][{"1752" 3...]]	7		
Other values (6600)		9669	
(Missing)	5743		
Toggle details			

<div> <div>VEP_CLIN_SIG</div> <div>Categorical</div> </div>			
Distinct count	77		
Unique (%)	1.1%		
Missing (%)	55.5%		
Missing (n)	8560		
uncertain_significance	1663		
uncertain_significance¬_provided	1472		
not_provided&pathogenic	845		
Other values (73)	2893		
(Missing)		8560	
Toggle details			

<div> <div>VEP_Codons</div> <div>Categorical</div> </div>			
Distinct count	5948		
Unique (%)	62.1%		
Missing (%)	37.9%		
Missing (n)	5848		
[[{"aAt/aGt" 7/9}][{"2/9"}]]	23		
[[{"Caa/Taa" 7/9}][{"2/9"}]]	20		
[[{"Gaa/Aaa" 7/9}][{"2/9"}]]	19		
Other values (5944)		9523	
(Missing)	5848		
Toggle details			

VEP_Consequence

Categorical

Distinct count	991		
Unique (%)	6.4%		
Missing (%)	0.0%		
Missing (n)	0		
[["missense_variant" "7/9"]][["upstream...		847	
[["missense_variant" "7/10"]][["non_co...		808	
[["intron_variant" "8/9"]][["intron_varian...		468	
Other values (988)			13310

Toggle details

VEP_DISTANCE

Categorical

Distinct count	7994		
Unique (%)	65.7%		
Missing (%)	21.2%		
Missing (n)	3270		
[[" "8/9"]][["2" "1/9"]]		12	
[[" "8/9"]][["5" "1/9"]]		9	
[[" "8/9"]][["768" "1/9"]]		9	
Other values (7990)			12133
(Missing)		3270	

Toggle details

VEP_EAS_MAF

Categorical

Distinct count	225		
Unique (%)	5.0%		
Missing (%)	71.1%		
Missing (n)	10967		
A:0-0000		859	
T:0-0000		838	
G:0-0000		718	
Other values (221)		2051	
(Missing)			10967

[Toggle details](#)

VEP_EA_MAF

Categorical

Distinct count	92
Unique (%)	7.4%
Missing (%)	91.9%
Missing (n)	14184

G:0	141
G:0-0001	135
T:0-0001	130
Other values (88)	843
(Missing)	14184

[Toggle details](#)

VEP_EUR_MAF

Categorical

Distinct count	264
Unique (%)	5.9%
Missing (%)	70.9%
Missing (n)	10939

A:0-0000	885
T:0-0000	825
G:0-0000	741
Other values (260)	2043
(Missing)	10939

[Toggle details](#)

VEP_EXON

Categorical

Distinct count	135
Unique (%)	1.3%
Missing (%)	34.8%
Missing (n)	5369

["11/27" "3/9"]	1550
-----------------	------

[[["11/27" "3/10"]][[" "2/10"]][["11/22" "...]]]	1010		
[[["10/27" "3/9"]][["10/28" "1/9"]][["10/10...]]]	572		
Other values (131)		6932	
(Missing)		5369	
Toggle details			

<div>VEP_ExAC_AFR_MAF</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>574</div> </div> <div> <div>Unique (%)</div> <div>12.5%</div> </div> <div> <div>Missing (%)</div> <div>70.4%</div> </div> <div> <div>Missing (n)</div> <div>10859</div> </div> <div> <div>G:0</div> <div>976</div> </div> <div> <div>C:0</div> <div>865</div> </div> <div> <div>T:0</div> <div>702</div> </div> <div> <div>Other values (570)</div> <div>2031</div> </div> <div> <div>(Missing)</div> <div>10859</div> </div>			
Toggle details			

<div>VEP_ExAC_AMR_MAF</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>449</div> </div> <div> <div>Unique (%)</div> <div>9.8%</div> </div> <div> <div>Missing (%)</div> <div>70.4%</div> </div> <div> <div>Missing (n)</div> <div>10862</div> </div> <div> <div>G:0</div> <div>1022</div> </div> <div> <div>C:0</div> <div>892</div> </div> <div> <div>A:0</div> <div>730</div> </div> <div> <div>Other values (445)</div> <div>1927</div> </div> <div> <div>(Missing)</div> <div>10862</div> </div>			
Toggle details			

<div>VEP_ExAC_Adj_MAF</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>1774</div> </div>			
---	--	--	--

Unique (%)	38.8%				
Missing (%)	70.3%				
Missing (n)	10856				
	T:8-24e-06	30			
	C:8-24e-06	29			
	A:8-237e-06	26			
	Other values (1770)	4492			
	(Missing)			10856	
Toggle details					

VEP_ExAC_EAS_MAF					
Categorical					
Distinct count	275				
Unique (%)	6.0%				
Missing (%)	70.3%				
Missing (n)	10857				
	G:0	1066			
	C:0	930			
	T:0	784			
	Other values (271)	1796			
	(Missing)			10857	
Toggle details					

VEP_ExAC_FIN_MAF					
Categorical					
Distinct count	161				
Unique (%)	3.6%				
Missing (%)	70.7%				
Missing (n)	10906				
	G:0	1165			
	C:0	1041			
	T:0	856			
	Other values (157)	1465			
	(Missing)			10906	
Toggle details					

VEP_ExAC_MAF

Categorical

Distinct count	690		
Unique (%)	15.1%		
Missing (%)	70.3%		
Missing (n)	10856		
	G:8-237e-06	226	
	C:8-237e-06	226	
	G:8-236e-06	219	
	Other values (686)	3906	
	(Missing)		10856

Toggle details

VEP_ExAC_NFE_MAF

Categorical

Distinct count	885		
Unique (%)	19.3%		
Missing (%)	70.3%		
Missing (n)	10857		
	G:0	569	
	C:0	537	
	A:0	402	
	Other values (881)	3068	
	(Missing)		10857

Toggle details

VEP_ExAC_OTH_MAF

Categorical

Distinct count	149		
Unique (%)	3.3%		
Missing (%)	70.4%		
Missing (n)	10861		
	G:0	1167	
	C:0	1028	
	A:0	849	
	Other values (145)	1528	

(Missing)

10861

Toggle details

VEP_ExAC_SAS_MAF

Categorical

Distinct count

509

Unique (%)

11.1%

Missing (%)

70.3%

Missing (n)

10857

G:0

970

C:0

833

A:0

690

Other values (505)

2083

(Missing)

10857

Toggle details

VEP_Existing_variation

Categorical

Distinct count

8619

Unique (%)

69.0%

Missing (%)

19.1%

Missing (n)

2944

rs80356834

7

rs28897718

7

rs397508986&CD982488&CX119479...

6

Other values (8615)

12469

(Missing)

2944

Toggle details

VEP_Feature

Categorical

Distinct count

89

Unique (%)

0.6%

Missing (%)

0.0%

Missing (n)

0

["XM_011535204-1"|"1/9"]

2297

[["ENST00000494123" "1/39"]][["ENST...		1518	
[["ENST00000530893" "1/10"]][["ENSE...		1027	
Other values (86)			10591
			Toggle details

VEP_Feature_type			
Categorical			
Distinct count	23		
Unique (%)	0.1%		
Missing (%)	0.0%		
Missing (n)	0		
Transcript			13411
[["Transcript" "38/39"]][["RegulatoryFe...		1518	
[["Transcript" "17/18"]][["RegulatoryFe...		151	
Other values (20)		353	
			Toggle details

VEP_GMAF			
Categorical			
Distinct count	582		
Unique (%)	12.9%		
Missing (%)	70.7%		
Missing (n)	10912		
A:0-0002		488	
G:0-0002		484	
T:0-0002		471	
Other values (578)		3078	
(Missing)			10912
			Toggle details

VEP_Gene			
Categorical			
Distinct count	80		
Unique (%)	0.5%		
Missing (%)	0.0%		

Missing (n)	0		
[[{"675" "4/9"}][{"ENSG00000139618" "...]	2297		
[[{"ENSG00000012048" "25/39"}][{"672...	1518		
[[{"ENSG00000139618" "4/10"}][{"675" ...	1027		
Other values (77)		10591	
Toggle details			

VEP_HGNC_ID		
Categorical		
Distinct count	68	
Unique (%)	0.4%	
Missing (%)	0.0%	
Missing (n)	0	
[[{"HGNC:1101" "7/9"}][{" "2/9"}]	3577	
[[{"HGNC:1100" "31/39"}][{" "8/39"}]	1518	
[[{"HGNC:1100" "31/38"}][{" "7/38"}]	1511	
Other values (65)		8827
Toggle details		

VEP_HGVSe
Constant
<i>This variable is constant and should be ignored for analysis</i>
Constant value

VEP_HGVSp
Constant
<i>This variable is constant and should be ignored for analysis</i>
Constant value

VEP_HIGH_INF_POS

Categorical

Distinct count	7
Unique (%)	87.5%
Missing (%)	99.9%

Missing (n)	15425		
	[[{"18/19"}][{"N"}"1/19"]]	2	
	[[{"10/12"}][{"N"}"2/12"]]	2	
	[[{"32/34"}][{"Y"}"2/34"]]	1	
	Other values (3)	3	
	(Missing)		15425
Toggle details			

VEP_IMPACT			
Categorical			
Distinct count	351		
Unique (%)	2.3%		
Missing (%)	0.0%		
Missing (n)	0		
	MODIFIER	5149	
	[[{"MODERATE"}"7/9"]][{"MODIFIER"}"2...]	1338	
	[[{"MODERATE"}"7/10"]][{"MODIFIER"}]"...]	852	
	Other values (348)		8094
Toggle details			

VEP_INTRON			
Categorical			
Distinct count	160		
Unique (%)	1.7%		
Missing (%)	39.2%		
Missing (n)	6045		
	[[{"25/39"}][{"9/21"}"3/39"]][{"10/21"}]"...]	1473	
	[[{"24/38"}][{"9/21"}"3/38"]][{"10/21"}]"...]	712	
	[[{"26/28"}][{"1/4"}"1/28"]][{"1/7"}"1/28"]]	266	
	Other values (156)		6937
	(Missing)		6045
Toggle details			

VEP_MOTIF_NAME			
Categorical			

Distinct count	7	
Unique (%)	87.5%	
Missing (%)	99.9%	
Missing (n)	15425	
[["" "24/25"] ["CTCF:MA0139-1" "1/25"]]	2	
[["" "10/12"] ["PU1:PB0058-1" "1/12"]]....	2	
[["" "12/14"] ["Cjun:MA0303-1" "2/14"]]	1	
Other values (3)	3	
(Missing)		15425

[Toggle details](#)

VEP_MOTIF_POS

Categorical

Distinct count	8	
Unique (%)	100.0%	
Missing (%)	99.9%	
Missing (n)	15425	
[["" "10/12"] ["1" "1/12"] ["2" "1/12"]]	2	
[["" "18/19"] ["2" "1/19"]]	1	
[["" "24/25"] ["13" "1/25"]]	1	
Other values (4)	4	
(Missing)		15425

[Toggle details](#)

VEP_MOTIF_SCORE_CHANGE

Categorical

Distinct count	8	
Unique (%)	100.0%	
Missing (%)	99.9%	
Missing (n)	15425	
[["" "10/12"] ["0-002" "1/12"] ["-0-002"...	2	
[["" "18/19"] ["-0-004" "1/19"]]	1	
[["" "32/34"] ["-0-163" "2/34"]]	1	
Other values (4)	4	
(Missing)		15425

[Toggle details](#)

VEP_PHENO

Categorical

Distinct count 20
Unique (%) 0.3%
Missing (%) 52.5%
Missing (n) 8110

1	4613
1&1	1444
1&1&1	698
Other values (16)	568
(Missing)	8110

[Toggle details](#)

VEP_PUBMED

Categorical

Distinct count 110
Unique (%) 35.9%
Missing (%) 98.0%
Missing (n) 15127

17916242	36
17719744	27
22684231	12
Other values (106)	231
(Missing)	15127

[Toggle details](#)

VEP_PolyPhen

Categorical

Distinct count 2438
Unique (%) 47.7%
Missing (%) 66.9%
Missing (n) 10317

[[{"probably_damaging(1)"|5/9"}][{"4/... 129

[[{"5/10"}]]["probably_damaging(1)"]...	53		
[[{"benign(0)"}]]["5/9"][[{"4/9"}]]	46		
Other values (2434)		4888	
(Missing)			10317
Toggle details			

<div> <div>VEP_Protein_position</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>4554</div> </div> <div> <div>Unique (%)</div> <div>47.0%</div> </div> <div> <div>Missing (%)</div> <div>37.2%</div> </div> <div> <div>Missing (n)</div> <div>5743</div> </div> </div>			
[[{"29/39"}]]["871"]["7/39"][[{"824"}]]["3/3..."]	11		
[[{"28/38"}]]["1086"]["7/38"][[{"1039"}]]["3..."]	11		
[[{"1782"}]]["7/9"][[{"2/9"}]]	10		
Other values (4550)			9658
(Missing)		5743	
Toggle details			

<div> <div>VEP_REFSEQ_MATCH</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>51</div> </div> <div> <div>Unique (%)</div> <div>0.3%</div> </div> <div> <div>Missing (%)</div> <div>0.0%</div> </div> <div> <div>Missing (n)</div> <div>0</div> </div> </div>			
[[{"5/9"}]]["rseq_mrna_match&rseq_e..."]		4046	
[[{"6/10"}]]["rseq_mrna_match&rseq_..."]	1839		
[[{"33/39"}]]["rseq_mrna_match&rseq_..."]	1670		
Other values (48)			7878
Toggle details			

<div> <div>VEP_SAS_MAF</div> <div>Categorical</div> <div> <div>Distinct count</div> <div>253</div> </div> <div> <div>Unique (%)</div> <div>5.7%</div> </div> </div>			
---	--	--	--

Missing (%)	71.1%				
Missing (n)	10967				
		A:0-0000	847		
		T:0-0000	804		
		G:0-0000	754		
		Other values (249)	2061		
		(Missing)		10967	
Toggle details					

VEP_SIFT					
Categorical					
Distinct count	1519				
Unique (%)	29.7%				
Missing (%)	66.9%				
Missing (n)	10320				
		[["deleterious(0)" "5/9"] [" "4/9"]]	293		
		[["5/10"] ["deleterious(0)" "5/10"]]	170		
		[["deleterious(0-01)" "5/9"] [" "4/9"]]	131		
		Other values (1515)	4519		
		(Missing)		10320	
Toggle details					

VEP_SOMATIC					
Categorical					
Distinct count	23				
Unique (%)	2.8%				
Missing (%)	94.6%				
Missing (n)	14604				
		0&1&1	337		
		0&0&1&1	124		
		1&1	110		
		Other values (19)	258		
		(Missing)		14604	
Toggle details					

VEP_STRAND

Categorical

Distinct count 43
Unique (%) 0.3%
Missing (%) 0.0%
Missing (n) 0

1	7658
-1	4120
["-1" "38/39"]["" "1/39"]	1518
Other values (40)	2137

Toggle details

VEP_SYMBOL

Categorical

Distinct count 68
Unique (%) 0.4%
Missing (%) 0.0%
Missing (n) 0

["BRCA2" "7/9"]["" "2/9"]	3577
["BRCA1" "31/39"]["" "8/39"]	1518
["BRCA1" "31/38"]["" "7/38"]	1511
Other values (65)	8827

Toggle details

VEP_SYMBOL_SOURCE

Categorical

Distinct count 66
Unique (%) 0.4%
Missing (%) 0.0%
Missing (n) 0

["EntrezGene" "4/9"]["HGNC" "3/9"][...	3309
["HGNC" "25/39"]["" "8/39"]["Entrez...	1518
["HGNC" "25/38"]["" "7/38"]["Entrez...	1511
Other values (63)	9095

Toggle details

VEP_TSL

Categorical

Distinct count 75**Unique (%)** 0.5%**Missing (%)** 0.0%**Missing (n)** 0

["6/9"]["1" "1/9"]["5" "1/9"]["2" "1/...	2863	
["15/39"]["1" "15/39"]["5" "6/39"][...	1518	
["6/10"]["1" "2/10"]["5" "1/10"]["2...	1039	
Other values (72)		10013

Toggle details

VEP_cDNA_position

Categorical

Distinct count 6899**Unique (%)** 68.6%**Missing (%)** 34.8%**Missing (n)** 5369

["23/39"]["2125" "3/39"]["1893" "2...	7	
["3190" "3/10"]["2957" "2/10"]["" "2/...	7	
["3/28"]["406" "3/28"]["1501" "2/2...	7	
Other values (6895)		10043
(Missing)	5369	

Toggle details

Variant_frequency(LOVD)

Categorical

Distinct count 80**Unique (%)** 0.5%**Missing (%)** 0.0%**Missing (n)** 0

-		12707
['-']	2273	
['1/1658']	143	
Other values (77)	310	

[Toggle details](#)

Variant_haplotype(LOVD)

Categorical

Distinct count 38

Unique (%) 0.2%

Missing (%) 0.0%

Missing (n) 0

-	12707
['-']	2690
['FA FANCD1_00038 -']	1
Other values (35)	35

[Toggle details](#)

polyPhen2_result(ESP)

Categorical

Distinct count 211

Unique (%) 1.4%

Missing (%) 0.0%

Missing (n) 0

-	14746
['None']	208
['None None None None None']	116
Other values (208)	363

[Toggle details](#)

Sample

	Source	Gene_symbol(ENIGMA)	Genomic_Coord
0	1000_Genomes	-	chr17:43073208:
1	ClinVar	-	chr17:43092777:

2	ENIGMA ClinVar	BRCA2	chr13:32387053:
3	1000_Genomes	-	chr17:43095356:
4	ENIGMA ClinVar 1000_Genomes	BRCA2	chr13:32353757:

In []: