In this assignment, we want to explore three genetic databases:

* UniProt
* dbSNP
* OMIM, …

Identify the entities (and name them) of each database and give an example.

Show how these databases are related.

OMIM

|  |  |
| --- | --- |
| **Entities** | **Attributes** |
| Disorder | OMIM ID  Disease name  Inheritance  Gene(s)  Protein number\*(foreign key from UniProt) |
| Gene | Gene MIM number  Gene name  Chromosomal location  Phenotype |
| Phenotype | Phenotype MIM number  Phenotype name  genes |

Example:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| OMIM ID | Disorder name | location | Gene name | Gene/Locus MIM number | Phenotype | Phenotype MIM number | Inheritance | Protein |
| 600185 | familial breast ovarian cancer | [17q21.31](https://www.omim.org/geneMap/17/603?start=-3&limit=10&highlight=603) | BRCA1 | [113705](https://www.omim.org/entry/113705) | {Breast-ovarian cancer, familial, 1} | [604370](https://www.omim.org/entry/604370) | AD, Mu | UniProt link |

UniProt

|  |  |
| --- | --- |
| **Entities** | **Attributes** |
| Protein | Protein number\* (primary key)  Protein name  Structure |
| Gene | Gene name  number of amino acids |
| Feature | rs number\* (foreign key from dbSNP)  disease variant  type of change |

Example:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Protein number | Protein name | Organism | Gene name | Number of AA | Structure | Disease variant | Rs number | Type of change |
| P51587 | Breast cancer type 2 susceptibility protein | Homo sapiens | BRCA2 | 3418 | Graphical information | VAR\_028167 25 | 80358961 | G->R |

dbSNP

|  |  |
| --- | --- |
| **Entities** | **Attributes** |
| SNP | Rs number\*(primary key)  Organism  position |
| Allele | Allele name  Allele frequency  Variation type |
| Variant | Sequence name  Change |

Example:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Rs number | Organism | Position | Allele variation type | Allele frequency | Sequence name | Change | Allele name |
| 80358961 | Homo sapiens | Chr:13:32319082 | snv | 0.000004 | LRG\_293 | NG\_012772.3:g.8603G>A | G>A/G>C/G>T |