**TBDR Technical Document**

**Assumptions**

* 1 variant can have many data sources and 1 data source can associate with multiple variants. (m : n)
* - 1 drug can have many variants and 1 variant can associate with many drugs. (m : n)

**ER Diagram**

**has**

1 m

**data\_sources**

**variants**

m 1

1 m mn

**has**

mn

**drug\_resistance**

m 1

mn

**drugs**

**DB Design**

Database: ***tbdr***

Table Name: **variants**

|  |  |  |
| --- | --- | --- |
| **PK** | id | INT (5) |
|  | var\_position\_genome\_start | INT (7) |
| **UK** | var\_position\_genome\_stop | INT (7) |
|  | var\_base | VARCHAR (6) |
|  | var\_type | VARCHAR (5) |
|  | number | INT (2) |
|  | wt\_base | VARCHAR (6) |
|  | region | VARCHAR (15) |
|  | gene\_id | VARCHAR (15) |
|  | gene\_name | VARCHAR (15) |
|  | gene\_start | INT (7) |
|  | gene\_stop | INT (7) |
|  | gene\_length | INT (4) |
|  | dir | VARCHAR (3) |
|  | wt\_aa | VARCHAR (5) |
|  | codon\_nr | INT (4) |
|  | codon\_nr\_e\_coli | INT (4) |
|  | var\_aa | VARCHAR (10) |
|  | aa\_change | VARCHAR (20) |
|  | codon\_change | VARCHAR (15) |
|  | var\_position\_gene\_start | INT (7) |
|  | var\_position\_gene\_stop | INT (7) |
|  | remarks | VARCHAR (1000) |

Table Name: **drugs**

|  |  |  |
| --- | --- | --- |
| **PK** | id | INT (4) |
|  | name | VARCHAR (30) |

Table Name: **data\_sources**

|  |  |  |
| --- | --- | --- |
| **PK** | id | INT (4) |
|  | name | VARCHAR (30) |
|  | description | VARCHAR (500) |

Table Name: **drug\_resistance** (maps m:n relationship)

|  |  |  |
| --- | --- | --- |
| **PK** | id | INT (5) |
| **FK** | variant\_id | INT (5) |
| **FK** | drug\_id | INT (4) |
| **FK** | data\_source\_id | INT (4) |
|  | reference\_pmid | INT (10) |
|  | high\_confidence | Boolean |