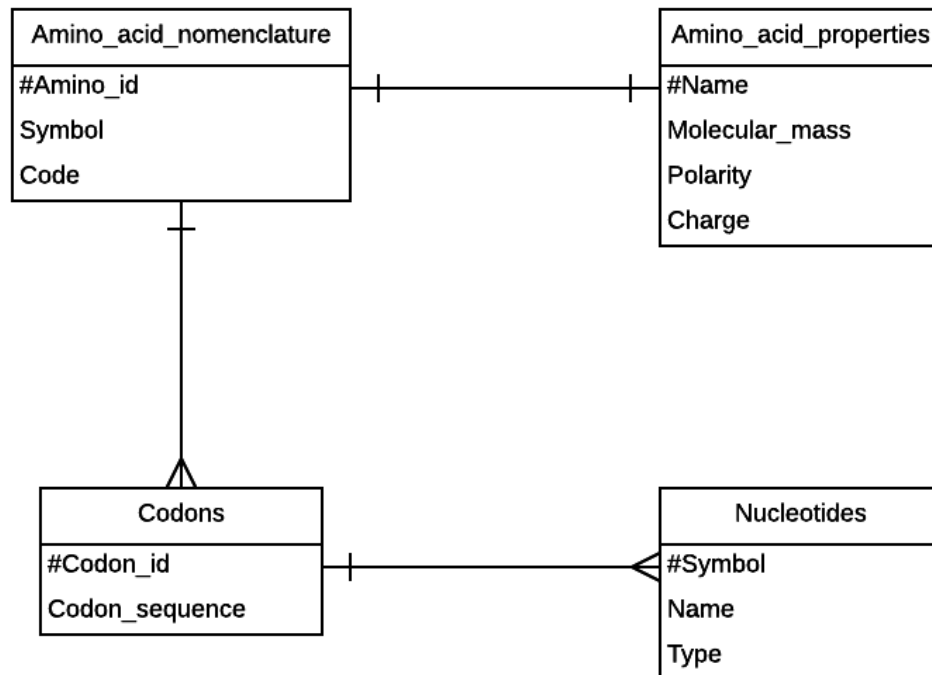


## INF115 – Assignment Two

1. One Nomenclature can have one set of properties, one Nomenclature can have many Codons and one Codon can have many Nucleotides.

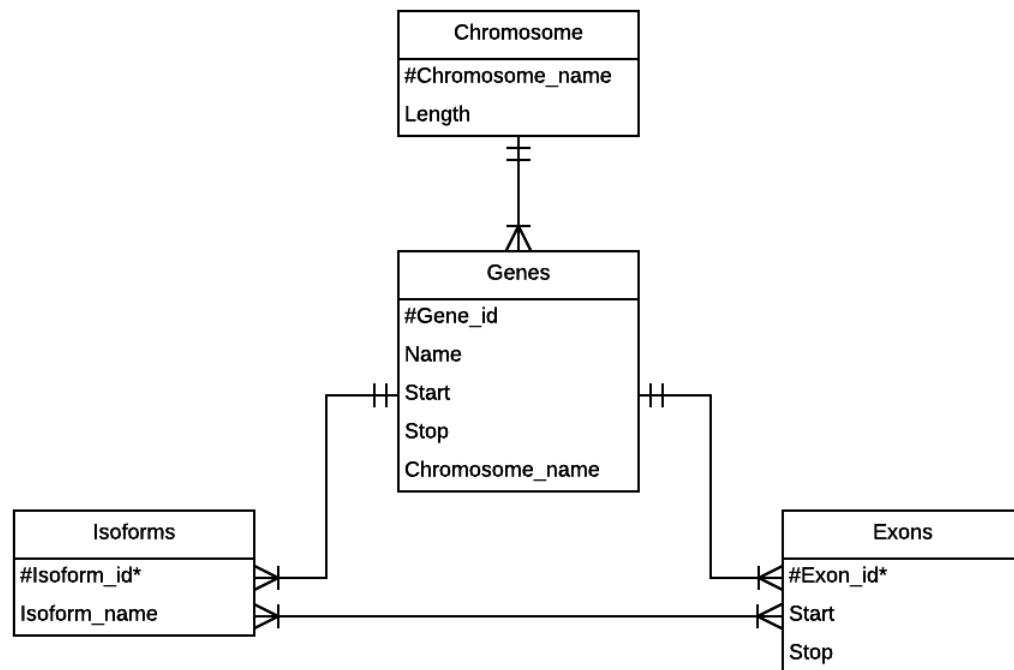


2.

1. Entities:

- Genes
- Exons
- Isoforms
- Chromosome

## 2. ER-Diagram:



## 3. DB in Third Normal Form:

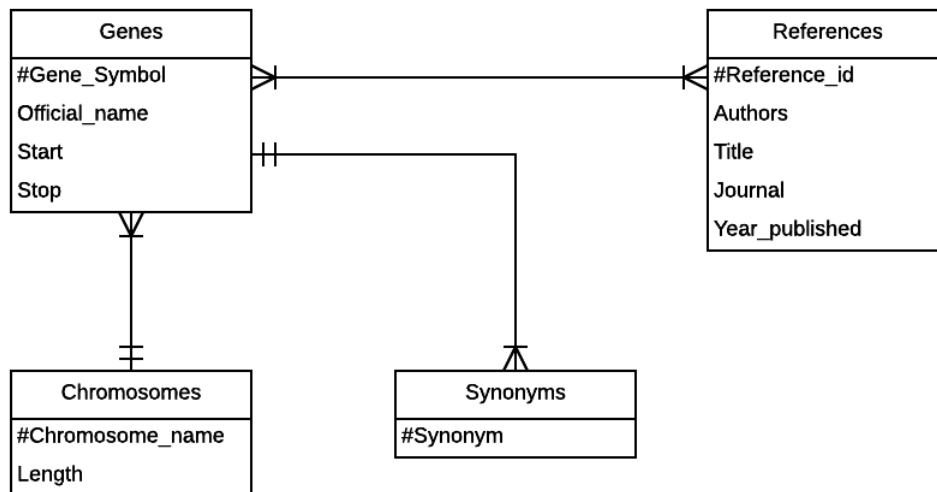
Chromosome(#Chromosome\_name, Length)  
 Genes(#Gene\_id, Name, Start, Stop, Chromosome\_name\*)  
 Exons(#Exon\_id, Start, Stop, Gene\_id\*)  
 Isoform(#Isoform\_id, Isoform\_name, Gene\_id\*)  
 Isoform\_exons(#Isoform\_id\*, #Exon\_id\*)

## 3.

### 1. Entities:

- Chromosomes
- Genes
- Synonyms
- References

## 2. ER-Diagram:



## 3. First Normal Form: Breaks second normal form since Chromosome\_Name is not dependent on Gene\_symbol.

Synonyms(#Synonym, Gene\_symbol\*)

Genes(#Gene\_symbol, Chromosome\_name, Chromosome\_length, Official\_name, Start\_coordinate, Stop\_coordinate, Reference\_id\*, Year\_published)

References(#Reference\_id, Authors, Title, Journal)

## 4.

Chromosome(#Chromosome\_name, Chromosome\_length)

Synonyms(#Synonym, #Gene\_symbol\*)

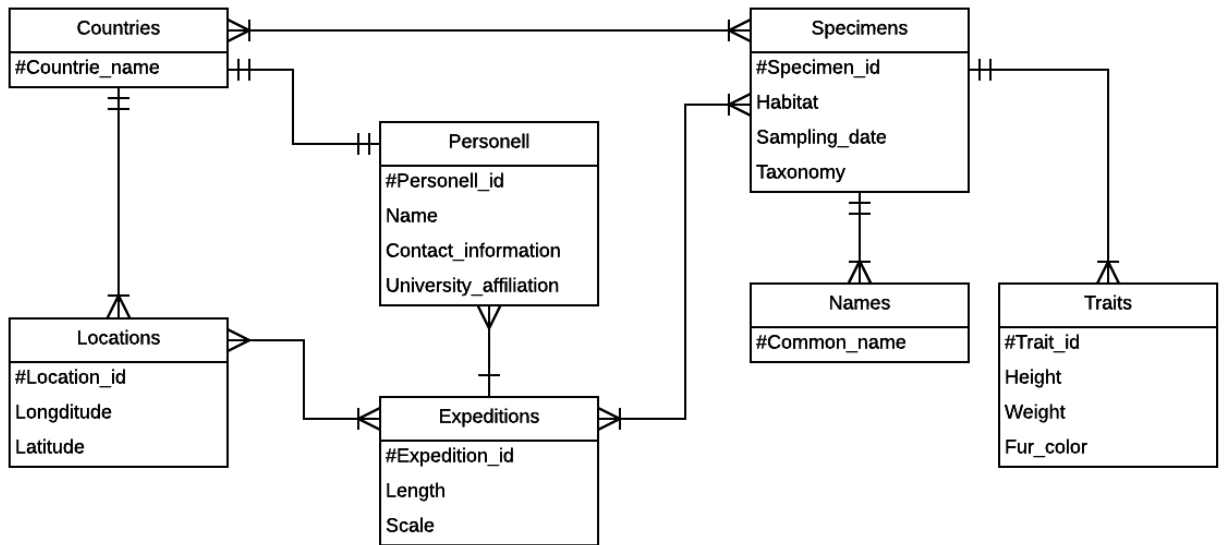
Gene(#Gene\_symbol, Chromosome\_name\*, Official\_name, Start\_coordinate, Stop\_coordinate)

GeneReference(#Gene\_symbol\*, #Reference\_id\*)

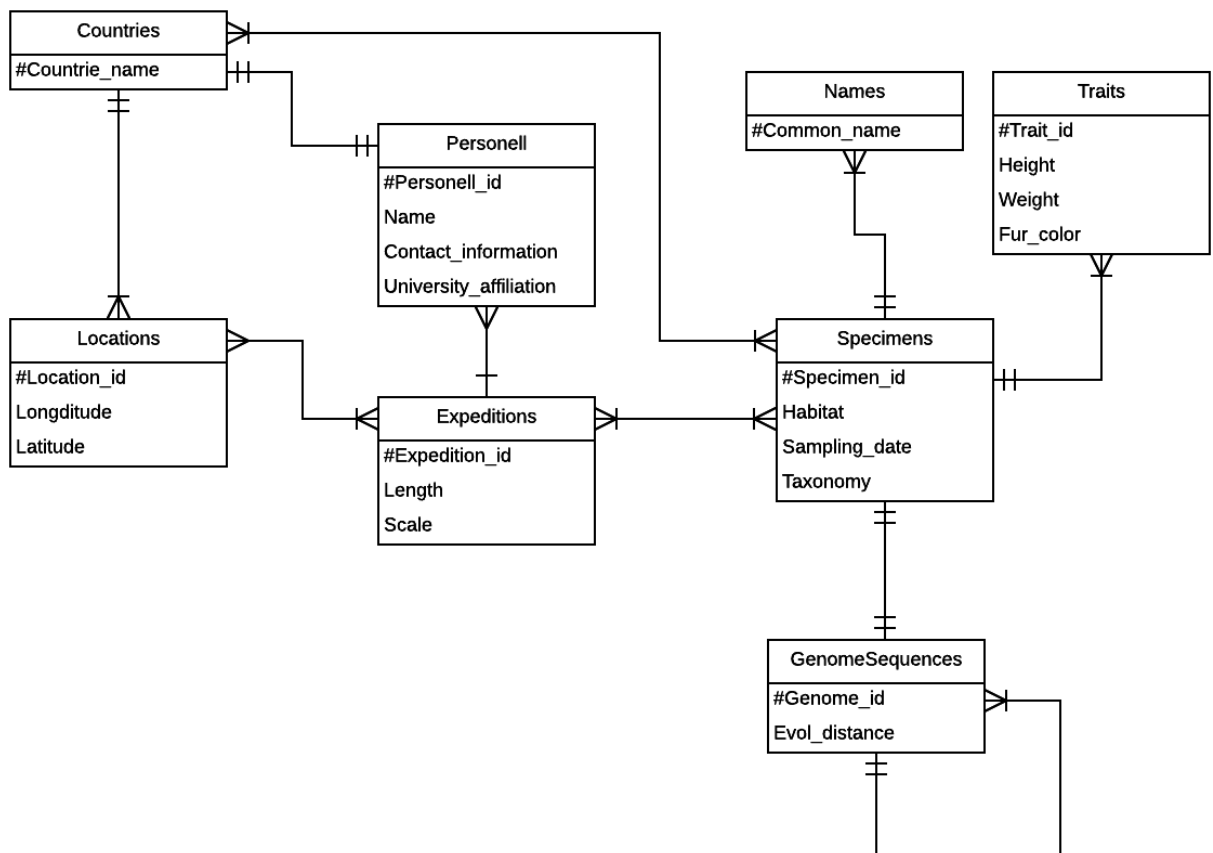
Reference(#Reference\_id, Authors, Title, Journal, Year\_published)

4.

1.



2. Two



5.

1. Errors

1. Dependency error in Patient: Where postcode is dependent on Region, not PatientID
2. Patient can only be in one location at one time so PatientToLocation is not needed.
3. The highest normalization level these tables upholds are 2NF

2. There will still be dependencies on dependencies. So BCNF is not fulfilled.

Here featureId will be dependent on OutbreakId and OutbreakID is dependent on GenomeID. There exists transitive dependencies. And the same value for OutbreakID and FeatureID will appear many times duplicates.

Not all values are directly dependent on the primary key

3. Functional Dependencies: x is dependent on y

1. OutbreakId > GenomeId
2. FeatureType > FeatureID

4. Candidate Keys: GenomeID, PatientID, OutbreakID, FeatureID

5. Normalization:

1. First Normal Form:

System (#PatientID, FirstName, SurName, Postcode, Region #SampleID, SampleDate, #TestID, TestName, ResistantToAntibiotic, #LocationID, LocationName, #GenomeID, OutbreakID, FeatureID, FeatureType, AntibioticName)

2. Second Normal Form

Patient(#PatientID, FirstName, SurName, Postcode, Region)

Sample(#SampleID, SampleDate, PatientID\*)

Labtest(#TestID, TestName, SampleID\*, ResistantToAntibiotic)

HospitalLocation(#LocationID, LocationName)

PatientToLocation(#PatientID\*, #LocationID\*)

GenomeSequence(#GenomeID, PatientID\*, OutbreakID, FeatureID, FeatureType, AntibioticName)

3. Third Normal Form

Patient(#PatientID, FirstName, SurName, Postcode\*)

Region(#Postcode, Region)

Sample(#SampleID, SampleDate, PatientID\*)

Labtest(#TestID, TestName, SampleID\*, ResistantToAntibiotic, OutbreakID)

HospitalLocation(#LocationID, LocationName)

PatientToLocation(#PatientID\*, #LocationID\*)

GenomeSequence(#GenomeID, PatientID\*, OutbreakID, FeatureID\*)

Feature(#FeatureID, FeatureType, AntibioticName)

4. BCNF

Same as Third Normal Form.