Homework #2: Entities

- We have 6 entity sets: gene, transcript, proteins, modified proteins, modified protein activity and modified protein localization (make sure there are at least 5 entity sets).

- The attributes used will include at least 3 each (they must at least cover the key attributes)
  - GENE (**Gene_ID**, **Gene_name**, MaximumCDS)
  - TRANSCRIPT (**Transcript_id**, **Gene _Name, Transcript_Sequence**)
  - PROTEIN (**Protein_ID**, Transcript_ID, Protein_sequence)
  - MOD_PROTEIN (**ModProtein_ID**, Protein_ID, ModProtein_Type)
  - MODPROTEIN_ACTIVITY (**ModProtein_ID** ,EC_No, Activity_Name,)
  - MODPROTEIN_LOCATION (**ModProtein_ID, Location_ID**, CellCycleStage)

Homework #2: Relationship Definitions

- Business rules will include the following
  - Each gene can produce zero to many transcripts
  - Each transcript comes from one and only one gene and produces exactly one protein
  - A protein sequence can be either measured or inferred.
  - A transcript can exist without a gene to which it maps
  - Each protein is produced by one and only one transcript.
  - A protein can exist without a transcript to which it maps.
  - A protein can have zero to many modified amino acids
  - A protein can be modified on more than one amino acid
  - A protein is a type of modified protein (type none) - recursive
  - Amino acids may be modified by more than one type of group.
  - A modified protein is existence-dependent on a protein
  - Proteins can be located at multiple locations in the cell

Homework #2

- Using MySQL Workbench
  - Create and ER IE model of the given entities
  - The model should include the relationships covering the rules defined above.
  - Due Thursday Feb 4 at 8:00 a.m.
  - Turn in the saved model file.