



Projet 1 : Base de Données de Gènes Humains

Contexte

Vous travaillez dans un laboratoire de génomique qui étudie les gènes humains impliqués dans différentes maladies. Le laboratoire a besoin d'un système simple pour organiser ses connaissances sur les gènes, leurs protéines correspondantes et les domaines fonctionnels de ces protéines.

Objectif

Créer une base de données relationnelle permettant de :

1. Stocker les informations sur les gènes humains
2. Associer chaque gène à ses protéines
3. Répertoire les domaines fonctionnels des protéines
4. Faciliter les recherches par caractéristiques

Entités à Modéliser

1. **Gène** : Unité d'information génétique
 - Symbole (ex: TP53)
 - Nom complet
 - Chromosome
 - Position sur le chromosome
 - Brin (+ ou -)
2. **Protéine** : Produit du gène
 - Numéro d'accèsion (ex: P04637)
 - Gène d'origine
 - Longueur en acides aminés
 - Masse moléculaire
 - Fonction biologique
3. **Domaine** : Région fonctionnelle de la protéine
 - Nom du domaine (ex: "Zinc finger")
 - Famille à laquelle il appartient
 - Description de sa fonction

Relations entre Entités



- Un **gène** peut coder pour une ou plusieurs **protéines**
- Une **protéine** est codée par un seul **gène**
- Une **protéine** peut contenir plusieurs **domaines**
- Un **domaine** peut être présent dans plusieurs **protéines**

Tâches à Réaliser

1. Créer un **MCD** (Modèle Conceptuel de Données) avec les entités et relations
2. Transformer le MCD en **MLD** (Modèle Logique de Données)
3. Définir les types de données appropriés pour chaque attribut
4. Écrire les scripts CREATE TABLE pour les différentes tables
5. Ajouter les contraintes nécessaires (Clés primaires, Clés étrangères, Contraintes CHECK)
6. Insérer au moins :
 - 15 gènes humains (ex: BRCA1, CFTR, HBB, etc.)
 - 20 protéines correspondantes
 - 10 domaines fonctionnels
 - 30 associations protéine-domaine
7. Écrire et exécuter 10 requêtes SQL

Exemple de requêtes :

1. Lister tous les gènes situés sur le chromosome 17
2. Trouver les protéines de plus de 500 acides aminés
3. Pour chaque protéine, compter le nombre de domaines qu'elle contient
4. Rechercher les gènes codant pour des protéines avec fonction "enzyme"
5. Identifier les domaines présents dans au moins 3 protéines différentes

Livrables Attendus

1. **Rapport** contenant :
 - MCD et MLD
 - Justification des choix de modélisation
 - Scripts SQL complets
 - Résultats des requêtes
2. **Fichiers SQL** :
 - Création des tables
 - Insertion des données
 - Requetes SQL