**The Protein Framework – sample queries**

1. **Identification of drugs targeting asthma biomarkers**

**(graph visualisation)**

**MATCH** ({DiseaseName:'ASTHMA'})<-[r1:BIOMARKER]-(protein:Protein)-[r2:DRUG\_TARGET]->(drug:Drug) **return** r1,r2

(table format)

**MATCH** ({DiseaseName:'ASTHMA'})<-[:BIOMARKER]-(protein:Protein)-[:DRUG\_TARGET]->(drug:Drug) **WITH** protein, collect(drug.DrugName) **AS** Drug\_Set **return** protein.UniprotId **as** Uniprot\_Id, Drug\_Set

2. **Return set of proteins, related to CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD) and ASTHMA, respectively, which have sequence similarity**:

**MATCH** ({DiseaseName: 'PULMONARY DISEASE, CHRONIC OBSTRUCTIVE'})<-[r1]-(protein1:Protein)-[r:SEQ\_SIM]->(protein2:Protein)–[r2]->({DiseaseName:'ASTHMA'}) **RETURN** r, r1,r2

3. **Identification of all shortest pathways of max. length = 3, (in terms of graph) between the Asthma node and other respiratory disease nodes (Chronic obstructive pulmonary disease, Hypertension essential):**

**MATCH** paths = allShortestPaths((disease1:Disease{DiseaseName: 'ASTHMA'})-[\*1..3]-(other:Disease)) **where** other.DiseaseName **in** ['PULMONARY DISEASE, CHRONIC OBSTRUCTIVE', 'HYPERTENSION, ESSENTIAL'] **RETURN** paths

4. **Identification of the shortest pathways (in terms of graph) between the CLOCK gene and Asthma:**

**MATCH** paths=shortestPath((protein:Protein{UniprotId:'O15516'})-[:PPI\_ASSOCIATION|BIOMARKER|SEQ\_SIM\*]-(d:Disease{DiseaseName:'ASTHMA'})) **RETURN** paths

5. **Identification of proteins that enhance the Lung tissue:**

**MATCH** ()-[r:TISSUE\_ENHANCED]->({TissueName:"LUNG"}) **RETURN** r **LIMIT** 25

6. **Set of proteins that are biomarkers for different diseases. The results show disease name, the protein list associated (based on UNIPROT ids) and the list size. Only the sets of proteins larger than 2 are considered.**

**MATCH** (protein:Protein)-[:BIOMARKER]->(disease)

**WITH** disease,

count(protein) **AS** Protein\_Set\_Size, collect(protein.UniprotId) **AS** Protein\_Set

**WHERE** Protein\_Set\_Size > 1

**RETURN** disease.DiseaseName **AS** DiseaseName, Protein\_Set, Protein\_Set\_Size

**ORDER** **BY** Protein\_Set\_Size **DESC**

7. **Similar for Drug-Protein Target association.**

**MATCH** (protein:Protein)-[: DRUG\_TARGET]->(drug)

**WITH** drug,

count(protein) **AS** Protein\_Set\_Size, collect(protein.UniprotId) **AS** Protein\_Set

**WHERE** Protein\_Set\_Size > 1

**RETURN** drug.DrugName **AS** DrugName, Protein\_Set, Protein\_Set\_Size

**ORDER** **BY** Protein\_Set\_Size **DESC**

8. **Identification of disease-neighbours, (in terms of graph), at distance < 6, between the Asthma node and other disease nodes:**

**MATCH** path = shortestPath((d:Disease{DiseaseName: 'ASTHMA'})-[\*1..5]-(other:Disease)) **RETURN**

**Distinct** other.DiseaseName **as** DiseaseNeighbour,

**CASE** **WHEN** LENGTH(path) <= 2 **THEN**

str('Neighbours')

**WHEN** LENGTH(path) <= 4 **THEN**

str('Neighbours of neighbours')

**ELSE**

str('Neighbours of neighbours of neighbours') **END** **AS** Relation, LENGTH(path) **as** Distance