SUPPLEMENTARY INFORMATION

Use of Graph Database for the Integration of Heterogeneous Biological Data

Byoung-Ha Yoon[1,2], Seon-Kyu Kim[1], Seon-Young Kim[1,2]*

[1]Personalized Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Korea,
[2]Department of Functional Genomics, University of Science and Technology (UST), Daejeon 34113, Korea

http://www.genominfo.org/src/sm/gni-15-19-s001.pdf.

Supplementary Fig. 2

3 layer search

  Mysql query
    SELECT A.DrugName, A.GeneSymbol, B.DiseaseName FROM Drug_Gene_Interaction A
    INNER JOIN Gene_Disease_Association B ON A.GeneSymbol = B.GeneSymbol
    WHERE A.role='increases^expression' and B.DiseaseName like '%cancer%';

  Neo4j Cypher query
    MATCH (drug)-[r:Drug_Gene_Interaction {role:'increases^expression'}]->
    (gene:Gene {GeneSymbol:'BRCA1'})-[r1]->(disease)
    WHERE disease.DiseaseName =~ '.*[Cc](ancer|ANCER).*'
    RETURN drug, r, gene, r1, disease

4 layer search

  Mysql query
    SELECT A.GeneName, B.ProteinName, C.PathwayName, D.DiseaseName FROM Gene_Protein_Interaction A,
    Protein_Disease_interaction B, Pathway_interaction C, Disease_associated D
    WHERE A.GeneName=B.GeneName and B.ProteinName=D.ProteinName and A.GeneName=C.GeneName and
    D.GeneName=A.GeneName and B.ProteinName=D.ProteinName and C.PathwayName='%SIGNALING%' and
    (D.DiseaseName='HYPERTENSION' or D.DiseaseName='ESSENTIAL');

  Neo4j Cypher query
    MATCH (gene)<- [r1:DEG_RELATED_TO]-(p:Protein)-[r2:PART_OF]->(d:Disease),(p:Protein)-
    [r3:IN_PATHWAY]->(path:Pathway), (p:Protein)-[r4:BIOMARKER]->(d:Disease)
    WHERE (d.DiseaseName IN ['HYPERTENSION','ESSENTIAL']) AND path.PathwayName=~'.*SIGNALING.*'
    AND (d.DiseaseName=~ '.*[Cc](ancer|ANCER).*') return r1,r2,r3,r4

Supplementary Figure 2. Cypher queries and MySQL queries to compare the search speed of Graph database and MySQL database.