Etapa 04 Projeto Banco de Dados

Tema: Análise em XQuery para exploração de anticorpos

Pedro Barros Bastos RA:204481
Gabriel Volpato Giliotti RA:197569
Rafael Cabral Pili RA:185999

Human Protein Atlas

Base: https://www.proteinatlas.org/ENSG00000134057.xml

```
▼<proteinAtlas xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://v19.proteinatlas.org/download/proteinatlas.xsd" schemaVersion="2.6">
 ▼<entry version="19" url="http://v19.proteinatlas.org/ENSG00000134057">
    <name>CCNB1</name>
    <synonym>CCNB</synonym>
   ▶<identifier id="ENSG00000134057" db="Ensembl" version="92.38">...</identifier>
   ▶<proteinClasses>...</proteinClasses>
   ▶<proteinEvidence evidence="Evidence at protein level">...</proteinEvidence>
   ▶<tissueExpression source="HPA" technology="IHC" assayType="tissue">...</tissueExpression>
   ▶<pathologyExpression source="HPA" technology="RNA" assayType="pathology">...</pathologyExpression>
   ▶<cellExpression source="HPA" technology="ICC/IF">...</cellExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="consensusTissue">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="tissue">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="humanBrainRegional">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="humanBrain">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="mouseBrainRegional">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="mouseBrain">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="pigBrainRegional">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="pigBrain">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="cellLine">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="blood">...</rnaExpression>
   ▶<rnaExpression source="HPA" technology="RNAseq" assayType="bloodLineage">...</rnaExpression>
   ▶<antibody id="CAB000115" releaseVersion="1.2" releaseDate="2006-03-13">...</antibody>
   ▶<antibody id="CAB003804" releaseVersion="2" releaseDate="2006-10-30" RRID="AB_562272">...</antibody>
   ▶<antibody id="HPA030741" releaseVersion="12" releaseDate="2013-12-05" RRID="AB_2673586">...</antibody>
   ▶<antibody id="HPA061448" releaseVersion="16" releaseDate="2016-12-04" RRID="AB_2684522">...</antibody>
  </entry>
 ▼<copyright>
    Copyrighted by the Human Protein Atlas, http://www.proteinatlas.org/about/licence
  </copyright>
</proteinAtlas>
```

Base a partir de sua amostragem hierárquica, partindo do documento em XML

- Entendimento da hierarquia da base: Descobrir proteínas que podem causar câncer, através da reação com anticorpos
- Problemas propostos:
  - Percentual de células cancerígenas dentre todos os anticorpos (total de tumores/total de amostras)
  - Percentual de células cancerígenas para cada anticorpo (total de tumores para cada anticorpo/total de amostras do anticorpo com tumores)

Análises e Resultados obtidos:

```
let $protein := doc('http://www.proteinatlas.org/ENSG00000134057.xml')

let $totalTissue := ($protein//proteinAtlas/entry/antibody/tissueExpression/data/tissueCell)
let $totalDeTecidos := count($totalTissue)

let $totalTumorTissue:=
($protein//proteinAtlas/entry/antibody/tissueExpression/data/tissueCell[contains(cellType/text(),'Tumor')])
let $totalDeTumores := count($totalTumorTissue)

for $c in ($protein//proteinAtlas/entry/antibody)
let $qtdTumoresPorAnticorpo := $c//tissueExpression/data/tissueCell[contains(cellType/text(),'Tumor')]
return count($qtdTumoresPorAnticorpo) div ($totalDeTecidos)*100

6.75675%

for $c in ($protein//proteinAtlas/entry/antibody)
let $qtdTumoresPorAnticorpo := $c//tissueExpression/data/tissueCell[contains(cellType/text(),'Tumor')]
return count($qtdTumoresPorAnticorpo) div ($totalDeTumores)*100

CAB000115   -->  33.3 %
CAB003804   -->  33.3%
HPA030741   -->  0.0%
HPA061448   -->  33.3%
```