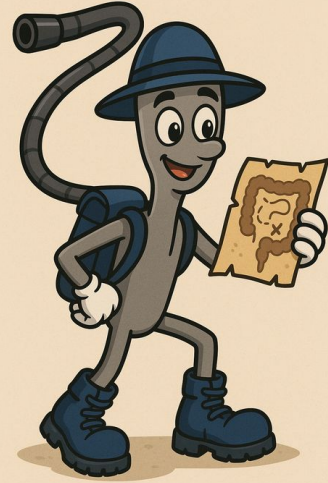


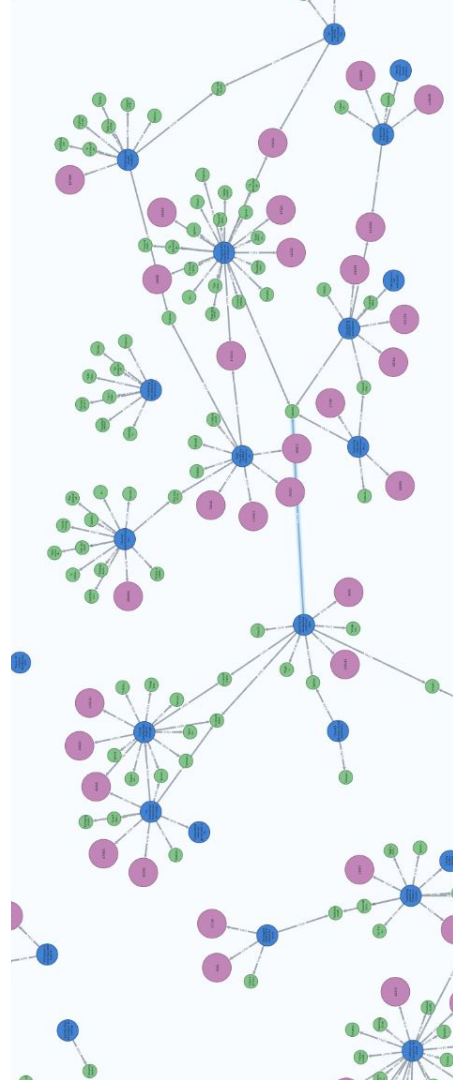
ECoGraph!



This ain't your grandma's genomic research.

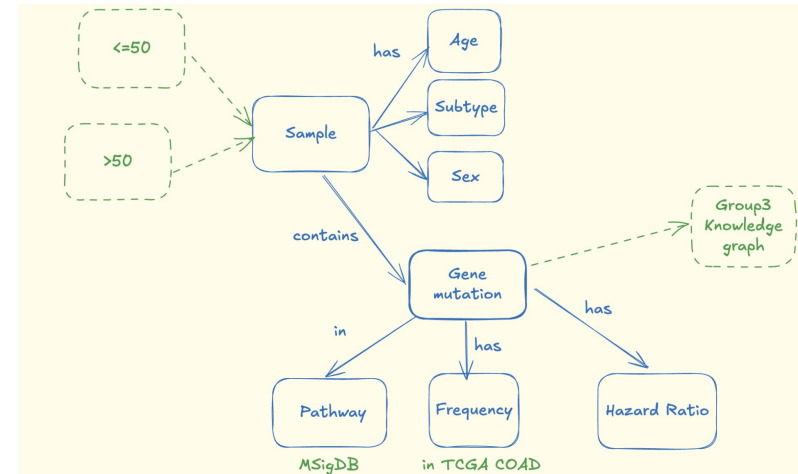
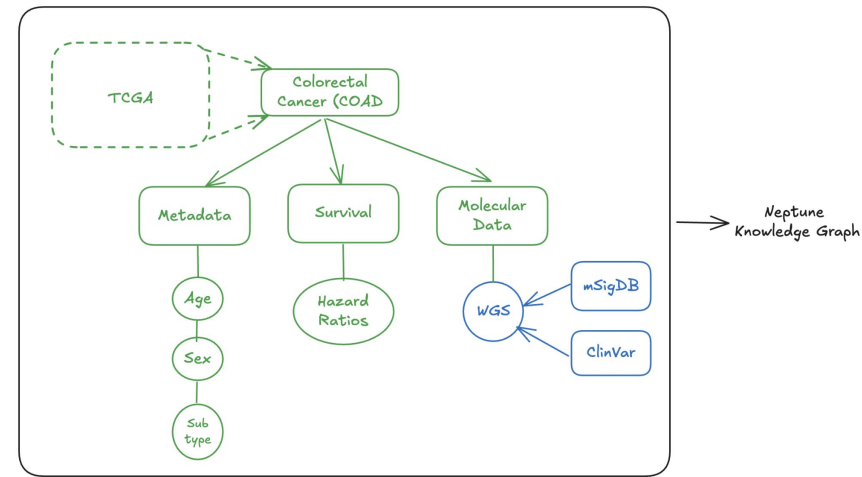
Background

- Colorectal adenocarcinoma (COAD) incidence has been **increasing** among younger people in the United States in recent decades
- Distribution of survival-relevant subtype and molecular patterns among early-onset COAD patients is not well understood
- EcoGraph integrates molecular and clinical data from colorectal cancer patients to **advance understanding of biomarkers, risk factors, and outcomes in early-onset colorectal cancer.**

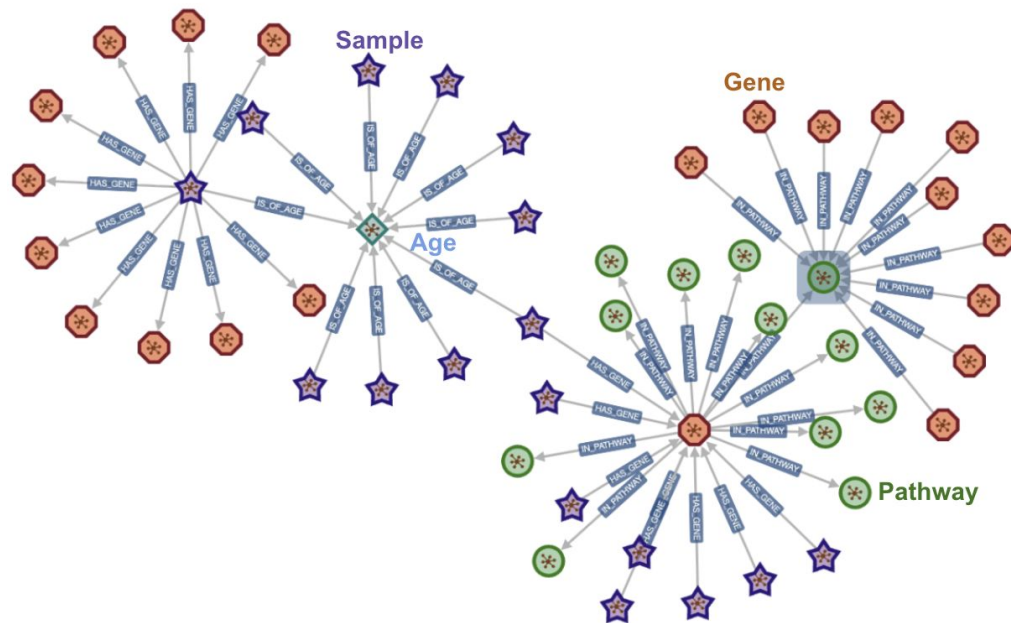


Materials and Methods

- Data: The Cancer Genome Atlas (TCGA)
- Calculated mortality hazard ratios (HRs) for association between gene mutation and age (≤ 50 years vs. >50 years) along with their interaction
- Mortality HRs per gene integrated with TCGA mutation data in Amazon Neptune to construct knowledge graph



Results: Overview of TCGA Knowledge Graph with Samples, Genes, Pathways, Subtypes of COAD and Clinical Features (Age)



- Nodes (# distinct values) defined as:
 - COAD Tumor sample (404)
 - Subtype (4)
 - Gene (21,333)
 - Pathway (2093)
 - Age (2; clinical feature)
- Edges (predicates) defined as:
 - Sample **HAS_GENE**
 - Sample **HAS_SUBTYPE**
 - Sample **IS_OF_AGE**
 - Gene **IN_PATHWAY**

Results

Of genes with a hazard ratio of nominal p-val of less than 0.005 what pathways have at least 2 genes in the list share

```
g.V().hasLabel('Gene').has('symbol',  
within("CUX1","SPARC","LAMC3","RNF17","PASD1","ATP2B3","PCDHA6","EXOC4","CDH6",  
,"KCNQ4","ARHGEF10","GRM4","CD248","MDM1")).
```

```
out('IN_PATHWAY').
```

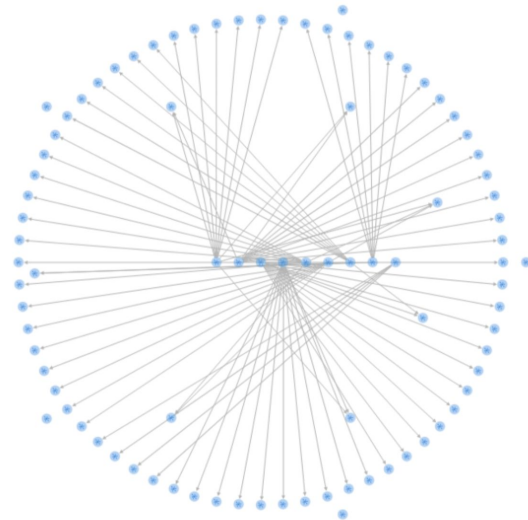
```
hasLabel('Pathway').
```

```
groupCount().
```

```
unfold().
```

```
filter(select(values).is(gte(2))).
```

```
select(keys)
```



Future directions

- Including Transcriptomic and Epigenomic data
- Integrating with CPTAC (Group 3) for proteomics
- Integrate with Bedrock (Group 9) to interpret knowledge graph

