Multi-omics data analysis in the cloud: Inference of differential breast cancer-related network hubs between TCGA patient cohorts

Kawther Abdillah1,3, Boris Aguilar2,3, Ronald C. Taylor4, George Acquaah-Mensah5

1General Dynamics Information Technology; 2Institute for Systems Biology; 3 Institute for Systems Biology-Cancer Genomics Cloud
4National Cancer Institute, Developmental Therapeutics Program, Rockville MD; 5Massachusetts College of Pharmacy and Health Sciences, SOP-Worcester

Figure 1

Breast invasive carcinoma (BrCA) remains a leading cause of mortality. Prognosis is worse for Black/African-American stage II BrCA patients 50 years old or younger (B/AA50) than for White patients of similar age and stage (W50).

Goal: To characterize the differences between these two cohorts using multi-omics data, including transcriptional regulatory relationships.

Methods

- TCGA data stored in Google BigQuery tables that are hosted by The Institute for Systems Biology - Cancer Genomics Cloud (ISB-CGC) were analyzed using standard SQL.
- All data analyses compared two BrCA cohorts (Black/African American women <=50 years old and White women <=50 years old).
- TCGA BigQuery tables containing information for the following data types were used for the analyses presented here: gene and microRNA expression, DNA methylation, and somatic mutation data.
- Network analysis was conducted which uses significant correlations between RNA-seq and miRNA data, which were generated using user-defined functions written in SQL.
- ISB-CGC TCGA BigQuery tables were accessed and queried in cloud-based R implementation using the bigRquery R package.
- Network differences and patterns were determined using the Reactome package in R.
- Functional overrepresentation analysis was conducted using the siggenes R package.
- Network differences and patterns were determined using the R methods package.

Figure 2

Summary and Conclusions

Cloud-based interrogations of a variety of BrCA and related data have tremendous potential. In this instance:

- Expressions of miR-6383, miR-548ag, and other cancer-regulating miR5s are statistically tied to that of master regulator HE4.
- Among genes expressed more in B/AA50 than W50, pathways associated with proliferation and the response to cellular stress are over-represented.
- Mutations of tumor suppressor CDH1 and PIK3CA occur more frequently and co-occur in W50.
- Mutations of HUWE1, HYDIN, and FBXW7 are more frequently mutated in B/AA50, and co-occur with mutations of TTN.

Figure 3

Figure 4

Figure 5a

Figure 5b

Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>B/AA50</th>
<th>W50</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBXW7</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SPO1</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>RNF213</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>CDH1</td>
<td>21</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PIK3CA</td>
<td>53</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>HYDIN</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>