Using Galaxy to analyze TCGA data: the experience of a Brazilian center

Presenter: Thais Talarico Hosokawa

Thais Talarico, Murilo Machado, Luciane Sussuchi da Silva, José Humberto Tavares Guerreiro Fregnani, Rui Manuel Reis, Adriane Feijó Evangelista
Barretos Cancer Hospital

- 3 branches in cancer: treatment, prevention, and Teaching and Researching
- Main unit – Barretos-SP
- IRCAD Latin America – Barretos-SP
- Mobile units factory – Barretos-SP
- Other units all over the country, non-profit

https://www.hcancerbarretos.com.br/
Molecular Oncology Research Center

- Grew fast because of an important funding
- Bioinformatics
  - 2 bioinformaticians (research(1), diagnosis(1))
  - 88 students (7 postdoctoral students, 15 PhD students, 23 Master students, 12 biomedical residents, 17 trainees, 10 others (undergraduate students, students with scholarship or visitors))
- 10 researchers
Three main barriers

- Biological background
- Infrastructure
- Skepticism about reliability
Our instance

- Some tools were not available on public servers -> to have admin privileges
- Huge amount of sensitive data;
- Better control of the processing according to our needs

Challenges

- Backgrounds mostly biological
- Troubleshooting -> documentation
A better and functional instance

- Active Interactive environment
- Using PLANEMO as tool wrapper to implement R scripts
Infrastructure

- We lost our instance (twice) in a shared server
- So, we decided to restart small in a dedicated server and as people get used to the interface, we would get a better structure
- We're gathering Galaxy's fans (specially from diagnosis)
- Support: David Clements and Björn Grüning, along with the community
Reliability

- People are resistant to changes, especially with a free software (less support/harder)
- During presentations and trainings, emphasize its advantages (culture)
- Wrapping algorithms that are used on diagnosis (and comparing results)
- Potential users suggested illustrative examples
The Cancer Genome Atlas - TCGA

- Cervical cancer data
- Time for approval not expected
- Analysing

About TCGA. https://cancergenome.nih.gov/abouttcga
Clinical data

- 2199 variables, tidy on the REDCap database, including file index
- API can be applied, but still not automated
Clinical data

<table>
<thead>
<tr>
<th>Histological type</th>
<th>count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervical squamous cell carcinoma</td>
<td>42</td>
<td>77.8</td>
</tr>
<tr>
<td>Mucinous adenocarcinoma of endocervical type</td>
<td>12</td>
<td>22.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HPV status</th>
<th>count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>6</td>
<td>11.1</td>
</tr>
<tr>
<td>Positive</td>
<td>48</td>
<td>88.9</td>
</tr>
</tbody>
</table>
Conclusion

✓ Some centers have no option unless using a local instance
✓ Have a good installation can be challenging
✓ Still, it’s an excellent option
✓ We got a functional instance and even if the infrastructure should be better before installing, we needed results to show
Acknowledgements

Advisors
Dr. Adriane Feijó Evangelista
Dr. Rui M.Reis

Post graduation at Barretos Cancer Hospital

DbGap process
Carolina Sgorlon Jorgetto (IAD)
Joyce Silva Pinto (EPIT)
Dr. Daniel Onofre Vidal
Daniela Girardi (COREME)
Marcella Marchioretto de Oliveira (Instituto Social)
Henrique Moraes Prata (Instituto Social)
Henrique Prata (Diretoria)

Other contributions
Dr. Fabiana de Lima Vazquez (CPOM)
Daniela Donadon (EPIT)
Maria Alice Dias (EPIT)
Cancer Registry Department

Installation
Dr. Adriane Feijó Evangelista
Dr. Luciane S.da Silva (CPOM)
Murilo Machado
Gabriel Andrade (TI)

Data analysis and interpreting
Dr. Adriane Feijó Evangelista
Dr. Rui M.Reis
Dr. José Humberto T.G.Fregnani
Dr. Luciane S.da Silva (CPOM)
Dr. Mariana Bisarro (CPOM)
Marcos Alves de Lima (NEB)

Barretos Cancer Hospital
Galaxy community
Thank you