Mining Human Prostate Cancer Datasets: The "camcAPP" Shiny App


Published in:
EBioMedicine

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2017 the authors. This is an open access article published under a Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
In Focus

Mining Human Prostate Cancer Datasets: The “camcAPP” Shiny App☆

Mark J. Dunning a, Sarah L. Vowler a, Emilie Lalonde b,c, Helen Ross-Adams d, Paul Boutros b,c, Ian G. Mills e, Andy G. Lynch a, Alastair D. Lamb a,f,g,*

Abstract

Obtaining access to robust, well-annotated human genomic datasets is an important step in demonstrating the relevance of experimental findings and, often, in generating the hypotheses that led to those experiments being conducted in the first place. We recently published data from the CamCaP Study Group which comprised two cohorts of men with prostate cancer who had undergone prostatectomy in Cambridge, UK and Stockholm, Sweden (Ross-Adams et al., 2015). We considered how we might best share our output with those who wish to interrogate the data with their own ideas, gene lists and clinical questions. We recognised that finding, downloading, pre-processing and assimilating any such dataset into a usable format is daunting and may put off many researchers. We also felt that interrogation tools generated to date (e.g. cBioPortal) lack functionality as they either cover too many organ types, or are limited in the extent, precision and tumour-site specificity of their clinical annotation. We therefore determined to produce an accessible web-based platform that would permit straightforward interrogation of these datasets with individual gene identifiers or gene sets. Furthermore, we decided to include additional ‘publicly-accessible’ human prostate cancer sets in order to increase the number of samples available and provide a degree of validation of any observations made across independent cohorts. We included a number of prominent publicly available sets with both gene expression and copy number data leading to a cohort of almost 500 men (Ross-Adams et al., 2015; Taylor et al., 2010; Grasso et al., 2012). We also included a small landmark series of expression data (Varambally et al., 2005). These studies are summarised in Table 1.

An important finding in our recent study was that prostate cancer could be divided into five distinct molecular subgroups based on stratification with a small number of copy number features which were also associated with RNA-expression change. These groups had different clinical outcomes. We wanted the app to allow researchers to determine the mean RNA-expression level or copy number status of a single gene or gene-set in prostatectomies from men divided either according to clinical categories (Gleason score, biochemical relapse status or tumour type) or according to molecular subgroups. These subgroups could either be pre-defined molecular groups published in the relevant papers, or de novo subgroups generated by hierarchical clustering based on an uploaded geneset.

We searched for other web-tools that are already available for this purpose. Although no such site exists for assessment of subgroup patterns or combined expression and copy number profiles, the Memorial Sloane Kettering Cancer Centre (MSKCC) and Michigan data (Table 1) can be analysed as part of cBioPortal (cBioPortal for Cancer Genomics, n.d.) along with the recently published prostate TCGA dataset (Robinson et al., 2015).

Here we introduce the camcAPP (http://bioinformatics.cruk.cam.ac.uk/apps/camcAPP/); a bespoke web interface to multiple prostate cancer genomics datasets. The interface was created with Shiny (https://www.rstudio.com/products/shiny/), and allows the non-specialist Bioinformatician to create publication-ready figures and tables through an intuitive interface to the underlying computer code.

After selecting a dataset of interest, and uploading a list of genes, the following analyses can be performed:

1) Boxplots and analysis of variance for expression of genes of interest grouped by clinical group, sample type, Gleason grade of copy-
Table 1: Summary of studies included in the cameAPP at initial release. Primary Tumours = tissue taken from radical prostatectomy specimens in men with clinical or metastatic disease. Advanced Tumours = tissue from channel transurethral resection of the prostate (chTURP) or prostatectomy in men with metastatic disease.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Paper</th>
<th>Platform: gene expression</th>
<th>Platform: copy number</th>
<th>Primary tumours</th>
<th>Advanced tumours</th>
<th>Clinical covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michigan 2005</td>
<td>Varambally et al. (2005)</td>
<td>Affymetrix U133 2.0</td>
<td>N/A</td>
<td>7</td>
<td>6</td>
<td>Sample Group</td>
</tr>
<tr>
<td>MSKCC 2010</td>
<td>Taylor et al. (2010)</td>
<td>Affymetrix Human 1.0 ST</td>
<td>Agilent 244k</td>
<td>109</td>
<td>19</td>
<td>Gleason, Copy Number Cluster</td>
</tr>
<tr>
<td>Michigan 2012</td>
<td>Grasso et al. (2012)</td>
<td>Agilent Whole Human 44k</td>
<td>Agilent 105k/244k</td>
<td>59</td>
<td>32</td>
<td>Sample Group</td>
</tr>
</tbody>
</table>

Study design: MD, AGL, ADL. CamCaP Study Group leads: ADL, HRA.
Data programming: MD, SV, AGL.
Programme contributions: EL.
Beta testing: IGM, EL, PB.
Oversight: AGL, ADL.

Disclosure
The authors have no conflicts of interest to declare.

Acknowledgements
The authors would like to thank Ola Bratt, Christof Kastner, Rajesh Nair and Declan Murphy for their critical and creative advice in review of the manuscript and/or design. We also acknowledge David Neal as the lead and principal investigator for the CamCaP Study Group.

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2017.02.022. This includes a ‘manual’ for the Shiny App which can also be downloaded from the app itself.

References
The authors would like to thank Ola Bratt, Christof Kastner, Rajesh Nair and Declan Murphy for their critical and creative advice in review of the manuscript and/or design. We also acknowledge David Neal as the lead and principal investigator for the CamCaP Study Group.

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2017.02.022. This includes a ‘manual’ for the Shiny App which can also be downloaded from the app itself.

References
- Nair and Declan Murphy for their critical and creative advice in review of the manuscript and/or design. We also acknowledge David Neal as the lead and principal investigator for the CamCaP Study Group.

Acknowledgements
The authors would like to thank Ola Bratt, Christof Kastner, Rajesh Nair and Declan Murphy for their critical and creative advice in review of the manuscript and/or design. We also acknowledge David Neal as the lead and principal investigator for the CamCaP Study Group.

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2017.02.022. This includes a ‘manual’ for the Shiny App which can also be downloaded from the app itself.

References
- Nair and Declan Murphy for their critical and creative advice in review of the manuscript and/or design. We also acknowledge David Neal as the lead and principal investigator for the CamCaP Study Group.

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2017.02.022. This includes a ‘manual’ for the Shiny App which can also be downloaded from the app itself.

References
- Nair and Declan Murphy for their critical and creative advice in review of the manuscript and/or design. We also acknowledge David Neal as the lead and principal investigator for the CamCaP Study Group.

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2017.02.022. This includes a ‘manual’ for the Shiny App which can also be downloaded from the app itself.

References