Event title	WORKSHOP: Working with genomics sequences and features in R with Bioconductor
Event type	Workshop
Date of event	23/09/2021
Time of event	10am-1pm AESST
Topic description	Explore the many useful functions that the Bioconductor environment offers for working with genomic data and other biological sequences. DNA and proteins are often represented as files containing strings of nucleic acids or amino acids. They are associated with text files that provide additional contextual information such as genome annotations. This workshop provides hands-on experience with tools, software and packages available in R via Bioconductor for manipulating, exploring and extracting information from biological sequences and annotation files. We will look at tools for working with some commonly used file formats including FASTA, GFF3, GTF, methods for identifying regions of interest, and easy methods for obtaining data packages such as genome assemblies.
Format description	Workshop, online via Zoom as described in https://zenodo.org/record/4158583 Paul Harrison led the training by introducing key concepts and demonstrating the steps involved in the analysis. Participants then moved into breakout rooms where they had the chance to apply these skills with support from facilitators. The workshop followed the tutorial linked in the 'Related work' section. A breakdown of timings and topics is provided in the schedule. Participation was free but subject to application with selection. Applications were reviewed by the organising committee.
Identifier(s)/URL	https://www.biocommons.org.au/events/bioconductor-genome
Licence	Materials are shared under a Creative Commons Attribution 4.0 International agreement unless otherwise stated on the materials
Keywords	R software Bioconductor Bioinformatics http://edamontology.org/topic_0091 Analysis http://edamontology.org/operation_2945

	Genomics http://edamontology.org/topic_0622 Sequence analysis http://edamontology.org/topic_0080
Contact	Melissa Burke (melissa@biocommons.org.au)
Audience	This workshop is open to Australian students and researchers who are using Bioconductor to analyse their life science data. Participants should have used a few Bioconductor packages, perhaps having worked through a tutorial vignette to perform a standard data analysis task. This workshop will assist those who are now interrogating their data in a way that requires going off the beaten path, or perhaps needing to better understand package vignettes mentioning things like "GRanges" and "TxDb".
Prerequisites	Some familiarity with R is assumed (prospective participants should already be comfortable with basic R concepts). It is not suitable for absolute beginners.
Technical requirements	Participants will need to have RStudio and R ready for use on their laptop.
Learning outcomes	After attending this workshop, participants will be able to: ● Work with data types in R representing DNA and amino-acid sequences, and genomic ranges (representing things such as genes, transcripts, exons, motif matches, or results of peak-calling). ● Perform useful operations on genomic ranges such as finding overlaps between two sets of ranges. ● Load and save data in commonly used file formats, appreciate some of the advantages and disadvantages of different formats. ● Install Bioconductor packages, find relevant documentation, and be aware of differences between the Bioconductor and base R ways of doing things.
Lead Trainer	Dr Paul Harrison, Monash Bioinformatics Platform
Facilitators	Dr Nandan Deshpande, Sydney Informatics Hub Ms Adele Barugahare, Monash Bioinformatics Platform Dr Andrew Perry, Monash Bioinformatics Platform Dr Nick Wong, Monash Bioinformatics Platform Mr Benjamin Reames, Monash Bioinformatics Platform
Related work	This workshop follows the tutorial ‘Working with DNA sequences and features in R with Bioconductor - version 2’ developed by This course has been developed for the Monash Bioinformatics Platform and Monash Data Fluency by Paul Harrison. https://monashdatafluency.github.io/r-bioc-2/