nf-core/circrna

A workflow for the quantification, miRNA target prediction and differential expression analysis of circRNAs

Barry Digby[1], Stephen P. Finn[2], Pilib Ó Broin[1]

[1] School of Mathematics, Statistics and Applied Mathematics, Áras de Brún, National University of Ireland, Galway
[2] Department of Histopathology and Morbid Anatomy, Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland

Introduction

Circular RNAs (circRNAs) are a class of non-coding RNAs formed by the back-splicing of precursor messenger RNA to create a covalently closed RNA loop structure (**figure 1**). circRNAs exhibit tissue- and stage-specific expression and are abundantly expressed in blood, serum and exosomes. Context dependent expression coupled with cytoplasmic stability make circRNAs ideal candidates as both diagnostic and prognostic biomarkers in a clinical setting, facilitating the use of non-invasive liquid biopsies to monitor disease status. Furthermore, circRNAs have been shown to harbour functionally active and evolutionarily conserved microRNA response elements (MREs), suggesting a regulatory role within the competing endogenous RNA (ceRNA) network **[1]**.
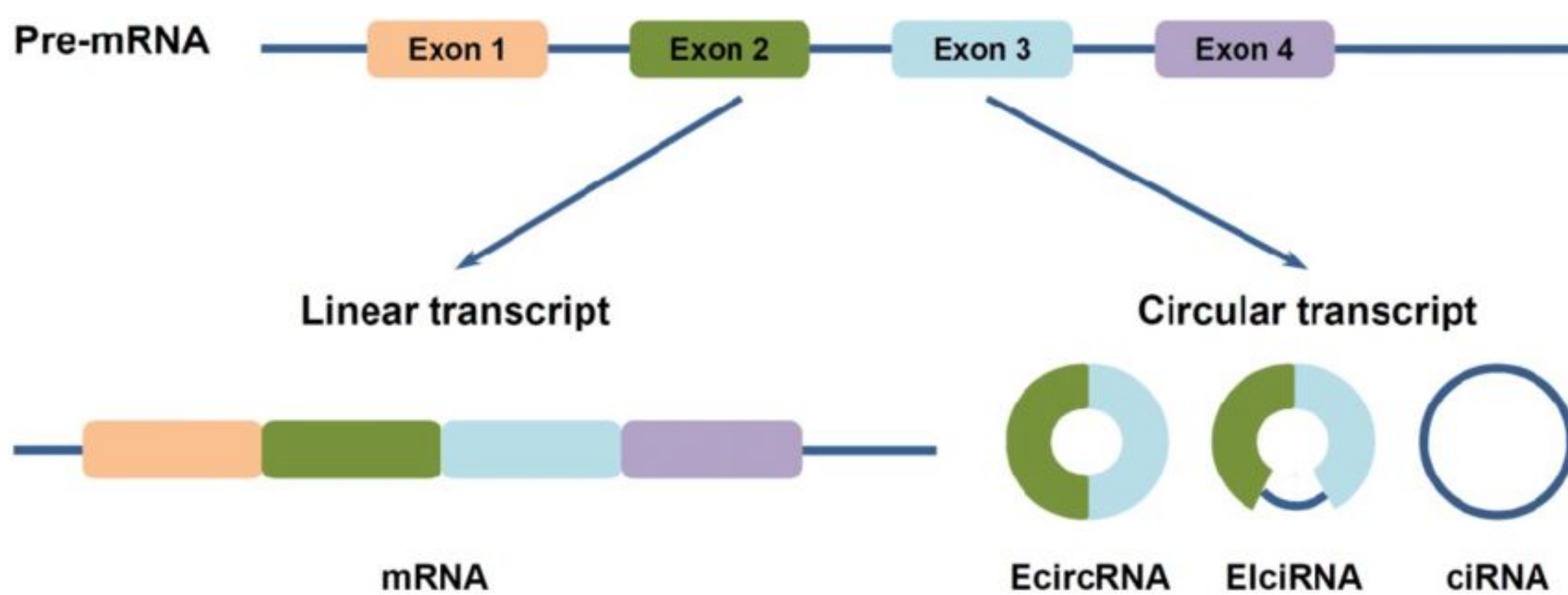
Figure 1: Backsplicing of pre-mRNA, producing circRNAs, ciRNAs & EI-ciRNAs **[2]**.

Current Landscape

Currently, there is no consensus on a circRNA quantification tool within the scientific community. The most popular tools utilise different read mappers and parsing strategies (de-novo, reference-guided) to detect non-canonical back-splice junction sites within sequencing reads. Only one tool is capable of performing differential expression tests between phenotypes of interest and no tool incorporates miRNA target prediction (**figure 2**).

To address these shortcomings, we present **nf-core/circrna**, a multi-functional, automated high-throughput pipeline implemented in Nextflow **[3]** that allows users to fully characterise the role of circRNAs in paired-end RNA-Seq datasets (**figure 3**).

Tool	Language	Mapper	De novo?	miRNA?	DEA?
CIRCexplorer2	Python	STAR/Tophat-fusion/BWA	yes	no	no
CIRIquant	Python	BWA/HISAT2	yes	no	yes
circRNA_finder	Perl	STAR	no	no	no
DCC	Python	STAR	no	no	no
find_circ	Python	Bowtie2	yes	no	no
MapSplice2	Python	Bowtie	no	no	no
Segemehl	C	Segemehl	yes	no	no

Figure 2: A brief summary of the installation requirements for several of the most popular circRNA quantification tools and their mode of operation.

Pipeline Workflow
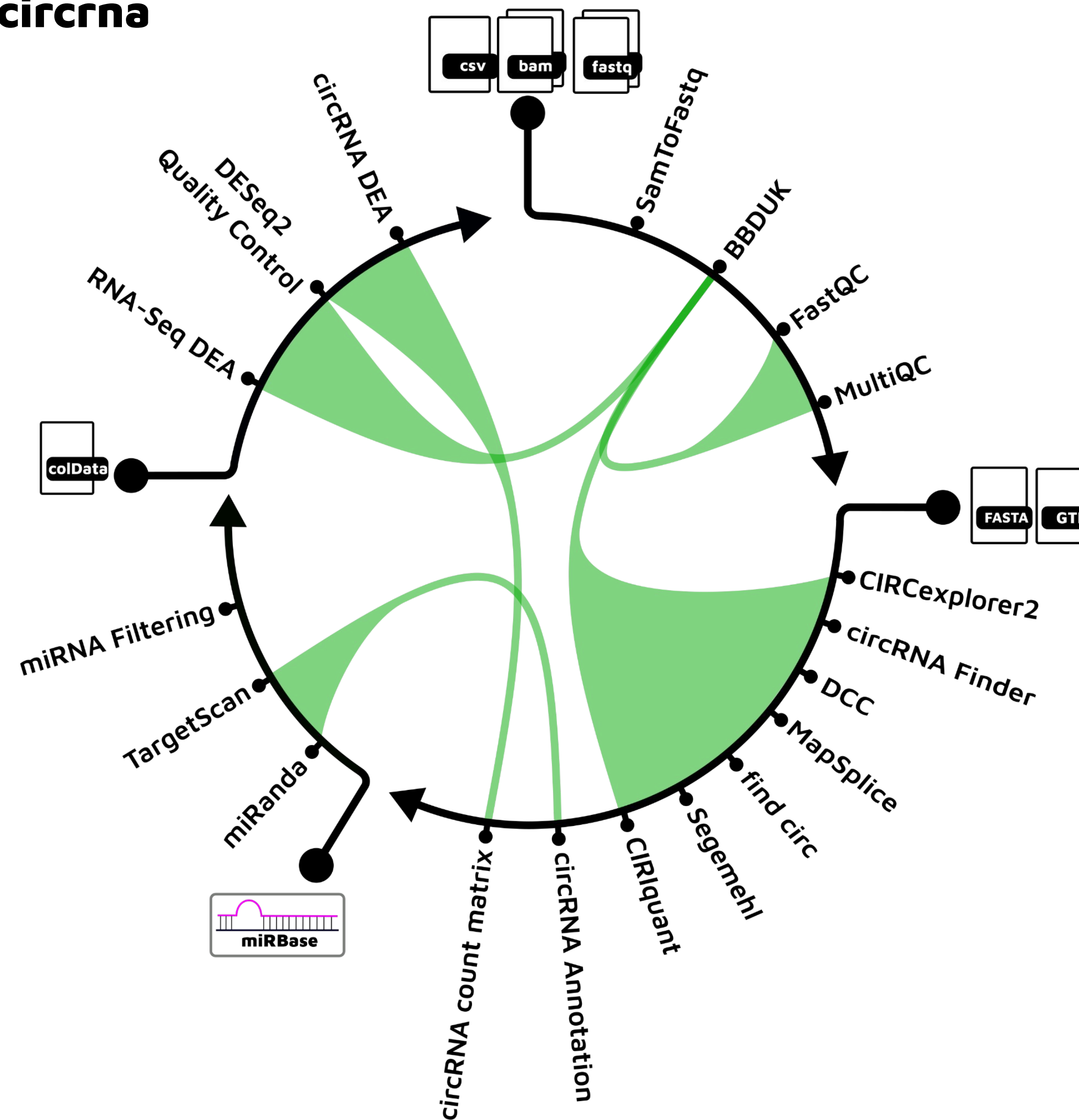
Figure 3: Schematic of nf-core/circrna workflow. The workflow is composed of 4 modules: **Pre-processing**, **circRNA Discovery**, **miRNA Prediction** & **Differential Expression Analysis**.

Availability

Source code, documentation and installation instructions are freely available at https://github.com/nf-core/circrna and https://nf-co.re/circrna. **nf-core/circrna** has been developed within the nf-core framework **[4]**, ensuring robust portability across POSIX compute environments, minimal installation requirements via containerisation using Docker/Singularity, compatibility with all HPC job schedulers and cloud based instances, comprehensive documentation and maintenance support. To run the workflow on a minimal test-dataset:

```
> nextflow pull nf-core/circrna
> nextflow run -r dev nf-core/circrna -profile test,<singularity/docker>
```

Contact

The authors of **nf-core/circrna** encourage engagement from the research community in the form of pull requests or issues via github. If you wish to contribute a novel or pre-existing circRNA tool to **nf-core/circrna**, please contact me at the usernames below to discuss integrating your tool within the workflow.

Selected Outputs
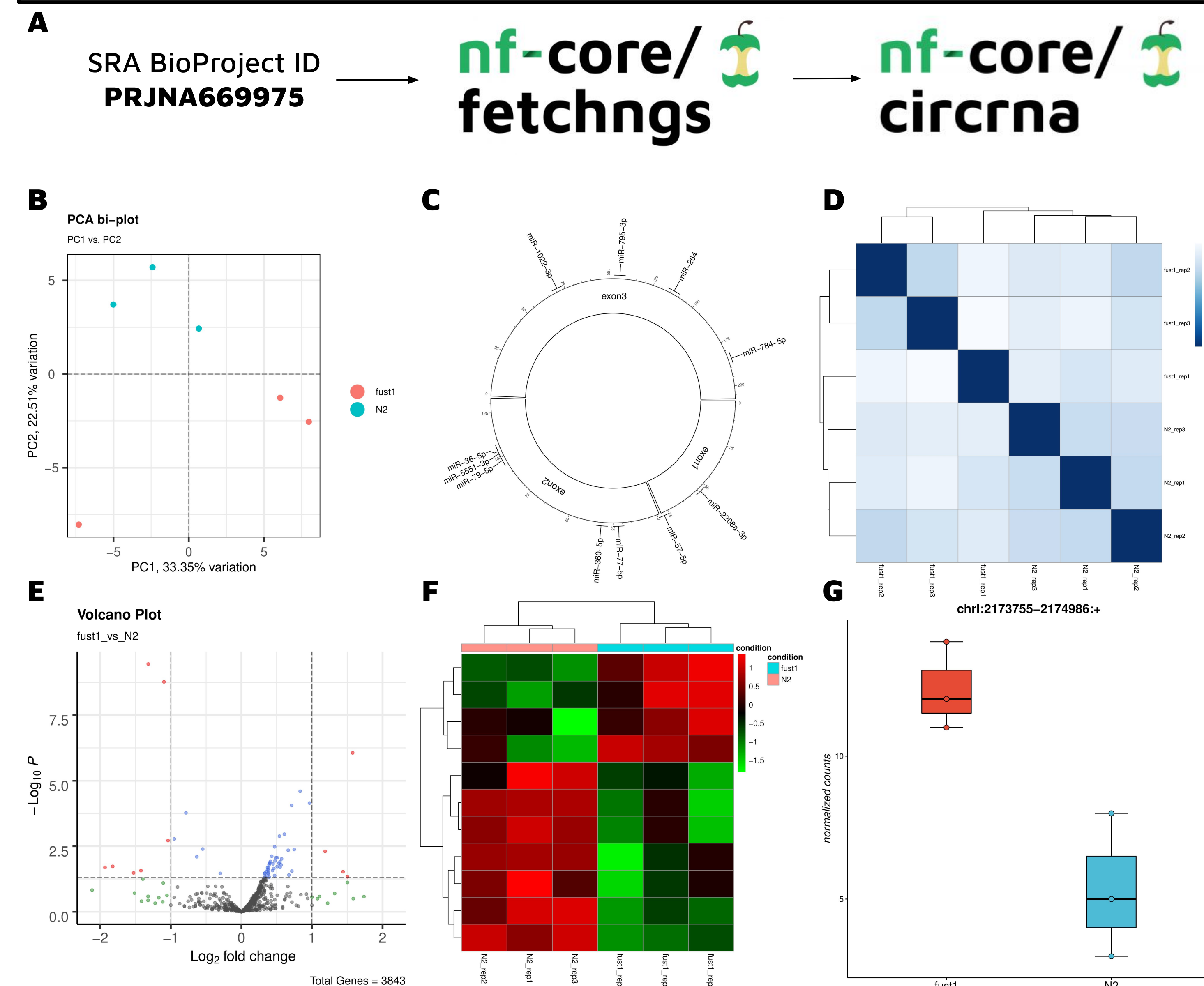
Figure 4: **A**: Demonstration of nf-core/circrna using BioProject accession number PRJNA669975 **[5]** in conjunction with nf-core/fetchngs **[6]**. **B**: PCA bi-plot of circRNA expression between *fust1* & *N2* strains. **C**: Circos plot of consensus miRNAs predicted by both miRanda and Targetscan within circRNA mature spliced sequence. **D**: Sample-to-sample clustering heatmap displaying sample heterogeneity. **E**: Volcano plot displaying differentially expressed circRNAs Log2FC values vs. p value (-log10). **F**: Heatmap of differentially expressed circRNAs. **E**: Normalised counts of differentially expressed circRNA chrl:2173755:2174986:+ between *fust1, N2* samples (N=3).

References

[1] Hansen, T. B., et al. (2013). Natural RNA circles function as efficient microRNA sponges. Nature, 495(7441), 384–388. https://doi.org/10.1038/nature11993
[2] Qiu LP et al. (2018) The Emerging Role of Circular RNAs in Hepatocellular Carcinoma. J Cancer; 9(9):1548-1559. https://www.jcancer.org/v09p1548.html
[3] Di Tommaso, P., et al. (2017) Nextflow enables reproducible computational workflows. Nat Biotechnol 35, 316–319, https://doi.org/10.1038/nbt.3820
[4] Ewels P, et al. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology 38:276-278, https://doi.org/10.1038/s41587-020-0439-xh
[5] D. Cao (2021) An autoregulation loop in fust-1 for circular RNA regulation in Caenorhabditis elegans. bioRxiv 2021.03.22.436400; doi: https://doi.org/10.1101/2021.03.22.436400
[6] nf-core/fetchngs (2021) https://doi.org/10.5281/zenodo.5070529

Follow @BarryDigby

BarryDigby