Coding & Best Practice in Programming: why it matters so much in the NGS era

Lex Nederbragt

Norwegian High-Throughput Sequencing Centre (NSC),Centre for Ecological and Evolutionary Synthesis (CEES), Dept. of Biosciences, University of Oslo, Oslo, Norway

Next generation sequencing has democratised large scale sequencing projects. No longer is generating a large sequencing dataset limited to genome centres. Neither is the analysis of large sequencing datasets limited to dedicated bioinformatics teams at genome centres or large institutions. Many more researchers now regularly obtain substantial NGS data for their projects. Wet lab biologists at all levels find themselves in need of acquiring computational skills to interpret their data. Coupled with a fast growing list of computational tools developed for the analysis of NGS data, this poses several challenges. Researchers inexperienced in judging computational tools need to choose the best/optimal one for the analysis of their data. Self-taught bioinformaticians are not familiar with best practices in computational science, leading to them make beginner's mistakes when they perform their computational analyses. Ultimately, this threatens correctness, reproducibility and reusability of the results obtained.

Luckily, all is not lost. There is an increasing awareness about the above issues, with reproducibility in computational science a topic receiving increased attention. Papers on best practices are published regularly. The non-profit organisation 'Software Carpentry' is helping out by organising two-day bootcamps where volunteers teach computational best practices to researchers.

In this talk, I will discuss what best practices in computational science are important, and why adhering to them, and teaching them, is crucial for our trust in the results obtained through NGS.