Currently there are many bioinformatics workflows in the nf-core. The parameters used by software of the workflows are mostly fixed, ensuring the reproducibility of the bioinformatics process. Moreover, nextflow can automatically record the version of the software used, which facilitates the reproduction of bioinformatics results if necessary.

Nextflow combining with container platforms improves the portability of bioinformatics workflows. Bioinformatics workflow, such as RNA-seq or Chip-seq, may contain tens of software packages, requiring a large amount of time to install them on local computer. However, nextflow supports container platforms such as Docker and Singularity, which means users only need to install Docker or Singularity before running workflows, enabling the entire workflows to deploy on different computers quickly. In other words, nextflow can greatly improve the portability of the bioinformatics workflow via working with container platforms.

Third, nextflow supports Amazon Web Service (AWS), which facilitates the processing of large-scale biological data. With the development of next-generation sequencing, the challenge is not to acquire high-throughput sequencing data, but to process data efficiently. The increasing amount of data places a growing demand on the computing power of local processors and storage space. For small labs or research institutions that do not focus on bioinformatics analysis, it may not be cost-effective to purchase high-performance computers (HPCs). However, the combination of nextflow and cloud computing will solve this dilemma. The AWS products used in our group's analysis include AWS Batch, AWS EC2 instance and AWS S3 bucket. With AWS Batch, nextflow can start EC2 spot instances with proper computing power for each process in the workflow. The computational power provided by a single EC2 instance is, in most cases, higher than that of a normal laptop or desktop computer. In addition, nextflow combining with AWS batch can start multiple instances to process multiple data sets at the same time, making it possible to process a large amount of genetic data in a short time. For example, using the same pipeline, it took three times longer to process six data sets locally (MacBook Air, M1, 16GB memory) than in EC2 spot instances. At the end of the sub-bio process, the spot instance is automatically terminated which means no extra costs will be generated. The final results can be automatically saved in AWS S3 bucket as a backup in the cloud.

In conclusion, nextflow together with AWS can process high-throughput biological data quickly and efficiently, and is highly portable and reproducible by working with container platform. All these features enable nextflow to become the mainstream trend of biological data processing in the feature.