nature methods

Focus on next-generation sequencing data analysis

A user's guide

An artistic interpretation of the sequencing process from the DNA molecule to the decoded bases by Erin Dewalt.

What used to take years and extensive collaborations—generating the raw sequence of the three gigabases in the human genome—can now be done in a few days by a single investigator using a single run on some of the latest next-generation sequencing machines. The drawback is that this massive amount of data comes in the form of short reads, and one needs to invest heavily in computational analysis and choose from a plethora of tools to make sense of it all.

The recurring theme when it comes to the choice of software is that a 'one-size-fits-all' program does not exist, but users have to mix and match, which requires knowledge about the analysis steps in a given application and how different software operates at each step. This Focus aims to guide readers in their choice of software so they can extract a maximum of information from the data.

Once the sequence reads are generated, the elementary step common to all applications is to align them to a reference or, if no reference is available, to assemble them *de novo*. Our first review explains the principle behind current alignment and assembly methods. The second review describes the different signatures created by structural variants and software suitable to detect each signature. Some datasets have unique features, such as those derived from chromatin immunoprecipitation (ChIP) or cDNA libraries, as in ChIP-sequencing (ChIP-seq) or RNA-sequencing (RNA-seq) experiments, respectively. These datasets require a different set of software, as discussed in the third review.

We realize that for some of these applications new algorithms are still emerging at a rapid rate. The goal of our authors was to explain the principles behind existing programs and to take the readers through the different analysis steps of an application so that they can make informed choices about software suitable for their needs.

We are pleased to acknowledge the financial support of Applied Biosystems. As always, *Nature Methods* carries sole responsibility for all editorial content and peer review.

Nicole Rusk

PUBLISHED ONLINE 15 OCTOBER 2009; DOI:10.1038/NMETH.F.271

CONTENTS

nature publishing group