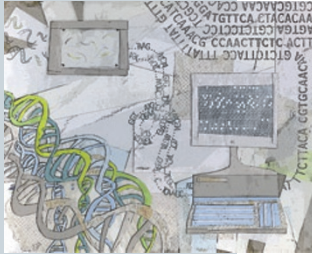


Focus on next-generation sequencing data analysis

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 provide some of this information.

Before choosing software, newcomers to next-generation sequencing will be faced with the choice of a platform, each with unique data characteristics. In this Focus the different sequencing technologies are discussed only peripherally; the main goal is to guide readers in their choice of software so they can extract a maximum of information from the data.

On the next two pages we introduce the authors who contributed to the Focus issue and provide a brief summary of the Commentary and the Reviews. Please visit our website for the full text.

We realize that for some of these applications new algorithms are still emerging at a rapid rate. The goal of our authors was not to give a comprehensive list of programs or to pit them against each other and declare a winner. Instead, in the Reviews the authors aim to explain the principles behind the programs and to take the readers through the different analysis steps of an application so that they can make informed choices about software suitable to their needs.

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

Nicole Rusk

The focus is freely available online thanks to the support of Applied Biosystems.
<http://www.nature.com/nmeth/journal/v6/n11s/index.html>



ONLINE CONTENTS

COMMENTARY

S2 Next-generation gap
John D McPherson

REVIEWS

S6 Sense from sequence reads: methods for alignment and assembly
Paul Flicek & Ewan Birney

S13 Computational methods for discovering structural variation with next-generation sequencing
Paul Medvedev, Monica Stanciu & Michael Brudno

S22 Computation for ChIP-seq and RNA-seq studies
Shirley Pepke, Barbara Wold & Ali Mortazavi

