

Preface

The human genome project was officially launched in 1990 equipped with a research funding of more than three billion USD. However, it took more than a decade and thousands of worldwide research institutes to discover and decode the full human genome sequence.

Nowadays, so-called next-generation sequencing devices process whole DNA and RNA within hours at moderate costs. Latest devices generate raw DNA reads with more than 30-times coverage in less than two days. However, interpretation and analysis of these raw data is still a time-consuming process potentially taking weeks. Next-generation sequencing devices are increasingly used in research and clinical environments to support treatment of specific diseases, such as cancer. This example highlights how fast the technological developments currently affect our daily lives.

Next-generation sequencing is also named to be the foundation for individual treatment decision, optimized therapies in course of personalized medicine and systems biology. Personalized medicine aims at treating patients specifically based on individual dispositions, such as genetic or environmental factors. However, the increasing amount of gathered diagnostic data requires specific software tools to identify relevant portions of data, process them at high-throughput, and provide ways to analyze them interactively.

We wrote this book to provide details about innovative approaches to process, combine, and analyze data required in the course of personalized treatment. It contains latest research results of applying in-memory database technology to process and analyze big genomic data. Furthermore, we share how to design and develop specific research tools that require real-time analysis of scientific data.

With this book, we contribute by bridging the gap between medical experts, such as physician, clinicians, and biological researchers, and technology experts, such as software developers, database specialists, and statisticians. As a result, we designed a specific structure of the book to support the individual audiences.

The book is structured as follows.

- Part I addresses the data acquisition, the modeling of processing and analysis pipelines, and how to accelerate preprocessing of data. This part is designed for bioinformaticians and researchers, who want to understand how to optimize the data preparation for their experiments.
- Part II gives examples how to design and implement specific applications enabling real-time analysis of scientific data. Furthermore, it provides guidelines to operate and to exchange huge data at fast pace. This part is intended for researchers and medical experts, who require to work with big data on a daily basis. It also provides guidelines for IT experts how to operate on these data from a software engineering perspective.

Potsdam, Oct 20, 2013

Hasso Plattner and Matthieu-P. Schapranow

High-Performance In-Memory Genome Data Analysis
How In-Memory Database Technology Accelerates
Personalized Medicine

Plattner, H.; Schapranow, M.-P. (Eds.)

2014, XXI, 223 p. 78 illus., Hardcover

ISBN: 978-3-319-03034-0