Big Data Analytics In Genomics Springer

Open Source Software in Life Science Research
Using Docker, GATK, and WDL in Terra
Developing a Cloud-based Secure Computation Platform for Genomics Research
Encyclopedia of Big Data Technologies
The Next Frontier for Innovation, Competition, and Productivity
Proceedings of CCB 2020
Big Data Analytics in Bioinformatics and Healthcare
Integrated Analysis and Causal Inference
Advances and Applications
Data Analysis for the Life Sciences with R
Machine Learning for Big Data Analysis
A Data Visualization Guide for Business Professionals
Contemporary Issues in Communication, Cloud and Big Data Analytics
Proceedings of CSI 2015
4th International Conference, BDA 2015, Hyderabad, India, December 15-18, 2015, Proceedings
High-Performance Big-Data Analytics
Proceeding of the International Conference on Computational and Bio Engineering, 2019, Volume 1
Contemporary Issues in Communication, Cloud and Big Data Analytics
Computing Systems and Approaches
Knowledge Modelling and Big Data Analytics in Healthcare
Advances in Computational and Bio-Engineering
Research Anthology on Big Data Analytics, Architectures, and Applications
BIG DATA ANALYTICS IN COMPUTATIONAL GENOME SEQUENCE ANALYSIS
Big Data Using Deep Learning Algorithms
Precision Public Health
Systems Analytics and Integration of Big Omics Data
Integrating Large-Scale Genomic Information into Clinical Practice
The Silent Cry: How to Turn Translational Medicine Towards Patients and Unmet Medical Needs
Gene Quantification
Big Data Management and Processing
Big Data in Radiation Oncology
Storytelling with Data
Big Data Analytics
Big Data Analytics in Genomics
Plant Breeding: Past, Present and Future
Mapping and Sequencing the Human Genome
An Introduction to Statistical Genetic Data Analysis
A Machine Learning Perspective

DOMINIQUE POTTS

Open Source Software in Life Science Research Springer
There is growing enthusiasm in the scientific community about the prospect of mapping and sequencing the human genome, a monumental project that will have far-reaching consequences for medicine, biology, technology, and other fields. But how will such an effort be organized and funded? How will we develop the new technologies that are needed? What new legal, social, and ethical questions will be raised? Mapping and Sequencing the Human Genome is a blueprint for this proposed project. The authors offer a highly readable explanation of the technical aspects of genetic mapping and sequencing, and they recommend specific interim and long-range research goals, organizational strategies, and funding levels. They also outline some of the legal and social questions that might arise and urge their early consideration by policymakers.
Using Docker, GATK, and WDL in Terra John Wiley & Sons
Genomics is a new and fast expanding area of biology encompassing high throughput or large scale experimentation at the whole genome level, and the organization, analysis and interpretation of the huge amount of data emerging from genome projects. Major new technologies have evolved recently that enable experimentation at the whole genome level, and more novel technologies are currently being developed. This volume describes in detail the new technology necessary to study the entire genome in a holistic manner and all the high throughput and large-scale experimental methodologies currently being used in genomic science. In addition the authors describe the progress of the newest technologies that are currently being developed. Written by experts in the field, this concise yet informative volume covers all aspects of technology pertaining to genomic studies. It is an essential book for anyone involved in genomic science.
Developing a Cloud-based Secure Computation Platform for Genomics Research Wiley-Scrivener
As technology evolves and electronic data becomes more complex, digital medical record management and analysis becomes a challenge. In order to discover patterns and make relevant predictions based on large data sets, researchers and medical professionals must find new methods to analyze and extract relevant health information. Big Data Analytics in Bioinformatics and Healthcare merges the fields of biology, technology, and medicine in order to present a comprehensive study on the emerging information processing applications necessary in the field of electronic medical record management. Complete with interdisciplinary research resources, this publication is an essential reference source for researchers, practitioners, and students interested in the fields of biological computation, database management, and health information technology, with a special focus on the methodologies and tools to manage massive and complex electronic information.
Encyclopedia of Big Data Technologies Bentham Science Publishers
Introduction to Molecular Genomics introduces the college student to the fundamental concepts of molecular biology and genomics. The text puts an emphasis on important topics in the subject that contribute to the learner’s understanding. These topics include molecular genomics, biodiversity and molecular phenomenon behind evolution of species, modern molecular methods for enhanced genomics research, DNA modifications at the molecular level for transgenic animal species, the role of cell environment on the gene expression, to name a few. The book has been designed to suit the requirements of educational courses in molecular biology, genomics and biochemistry. Key features - Covers basic concepts on key topics in molecular biology and genomics - Simple easy-to-read layout - Includes references for further reading - Includes a section on ethical aspects of scientific research Introduction to Molecular Genomics is a simple primer for students in applied or advanced life science courses at undergraduate levels
The Next Frontier for Innovation, Competition, and Productivity Springer
Knowledge Modelling and Big Data Analytics in Healthcare: Advances and Applications focuses on automated analytical techniques for healthcare applications used to extract knowledge from a vast amount of data. It brings together a variety of different aspects of the healthcare system and aids in the decision-making processes for healthcare professionals. The editors connect four contemporary areas of research rarely brought together in one book: artificial intelligence, big data analytics, knowledge modelling, and healthcare. They present state-of-the-art research from the healthcare sector, including research on medical imaging, healthcare analysis, and the applications of artificial intelligence in drug discovery. This book is intended for data scientists, academicians, and industry professionals in the healthcare sector.
Proceedings of CCB 2020 Frontiers Media SA
Through this book, researchers and students will learn to use R for analysis of large-scale genomic data and how to create routines to automate analytical steps. The philosophy behind the book is to start with real world raw datasets and perform all the analytical steps needed to reach final results. Though theory plays an important role, this is a practical book for graduate and undergraduate courses in bioinformatics and genomic analysis or for use in lab sessions. How to handle and manage high-throughput genomic data, create automated workflows and speed up analyses in R is also taught. A wide range of R packages useful for working with genomic data are illustrated with practical examples. The key topics covered are association studies, genomic prediction, estimation of population genetic parameters and diversity, gene expression analysis, functional annotation of results using publically available databases and how to work efficiently in R with large genomic datasets. Important principles are demonstrated and illustrated through engaging examples which invite the reader to work with the provided datasets. Some methods that are discussed in this volume include: signatures of selection, population parameters (LD, FST, FIS, etc); use of a genomic relationship matrix for population diversity studies; use of SNP data for parentage testing; snpBLUP and gBLUP for genomic prediction. Step-by-step, all the R code required for a genome-wide association study is shown: starting from raw SNP data, how to build databases to handle and manage the data, quality control and filtering measures, association testing and evaluation of results, through to identification and functional

annotation of candidate genes. Similarly, gene expression analyses are shown using microarray and RNAseq data. At a time when genomic data is decidedly big, the skills from this book are critical. In recent years R has become the de facto tool for analysis of gene expression data, in addition to its prominent role in analysis of genomic data. Benefits to using R include the integrated development environment for analysis, flexibility and control of the analytic workflow. Included topics are core components of advanced undergraduate and graduate classes in bioinformatics, genomics and statistical genetics. This book is also designed to be used by students in computer science and statistics who want to learn the practical aspects of genomic analysis without delving into algorithmic details. The datasets used throughout the book may be downloaded from the publisher's website.

Big Data Analytics in Bioinformatics and Healthcare Springer

This book presents the outcomes of the First International Conference on Communication, Cloud, and Big Data (CCB) held on December 18–19, 2020, at Sikkim Manipal Institute of Technology, Majitar, Sikkim, India. This book contains research papers and articles in the latest topics related to the fields like communication networks, cloud computing, big data analytics, and on various computing techniques. Research papers addressing security issues in above-mentioned areas are also included in the book. The research papers and articles discuss latest issues in the above-mentioned topics. The book is very much helpful and useful for the researchers, engineers, practitioners, research students, and interested readers.

Integrated Analysis and Causal Inference Big Data Analytics in Genomics

This book gathers state-of-the-art research in computational engineering and bioengineering to facilitate knowledge exchange between various scientific communities. Computational engineering (CE) is a relatively new discipline that addresses the development and application of computational models and simulations often coupled with high-performance computing to solve complex physical problems arising in engineering analysis and design in the context of natural phenomena. Bioengineering (BE) is an important aspect of computational biology, which aims to develop and use efficient algorithms, data structures, and visualization and communication tools to model biological systems. Today, engineering approaches are essential for biologists, enabling them to analyse complex physiological processes, as well as for the pharmaceutical industry to support drug discovery and development programmes.

Advances and Applications Horizon Scientific Press

Big Data in Omics and Imaging: Integrated Analysis and Causal Inference addresses the recent development of integrated genomic, epigenomic and imaging data analysis and causal inference in big data era. Despite significant progress in dissecting the genetic architecture of complex diseases by genome-wide association studies (GWAS), genome-wide expression studies (GWES), and epigenome-wide association studies (EWAS), the overall contribution of the new identified genetic variants is small and a large fraction of genetic variants is still hidden. Understanding the etiology and causal chain of mechanism underlying complex diseases remains elusive. It is time to bring big data, machine learning and causal revolution to developing a new generation of genetic analysis for shifting the current paradigm of genetic analysis from shallow association analysis to deep causal inference and from genetic analysis alone to integrated omics and imaging data analysis for unraveling the mechanism of complex diseases. FEATURES Provides a natural extension and companion volume to Big Data in Omic and Imaging: Association Analysis, but can be read independently. Introduce causal inference theory to genomic, epigenomic and imaging data analysis Develop novel statistics for genome-wide causation studies and epigenome-wide causation studies. Bridge the gap between the traditional association analysis and modern causation analysis Use combinatorial optimization methods and various causal models as a general framework for inferring multilevel omic and image causal networks Present statistical methods and computational algorithms for searching causal paths from genetic variant to disease Develop causal machine learning methods integrating causal inference and machine learning Develop statistics for testing significant difference in directed edge, path, and graphs, and for assessing causal relationships between two networks The book is designed for graduate students and researchers in genomics, epigenomics, medical image, bioinformatics, and data science. Topics covered are: mathematical formulation of causal inference, information geometry for causal inference, topology group and Haar measure, additive noise models, distance correlation, multivariate causal inference and causal networks, dynamic causal networks, multivariate and functional structural equation models, mixed structural equation models, causal inference with confounders, integer programming, deep learning and differential equations for wearable computing, genetic analysis of function-valued traits, RNA-seq data analysis, causal networks for genetic methylation analysis, gene expression and methylation deconvolution, cell –specific causal networks, deep learning for image segmentation and image analysis, imaging and genomic data analysis, integrated multilevel causal genomic, epigenomic and imaging data analysis.

Data Analysis for the Life Sciences with R Springer

A comprehensive introduction to modern applied statistical genetic data analysis, accessible to those without a background in molecular biology or genetics. Human genetic research is now relevant beyond biology, epidemiology, and the medical sciences, with applications in such fields as psychology, psychiatry, statistics, demography, sociology, and economics. With advances in computing power, the availability of data, and new techniques, it is now possible to integrate large-scale molecular genetic information into research across a broad range of topics. This book offers the first comprehensive introduction to modern applied statistical genetic data analysis that covers theory, data preparation, and analysis of molecular genetic data, with hands-on computer exercises. It is accessible to students and researchers in any empirically oriented medical, biological, or social science discipline; a background in molecular biology or genetics is not required. The book first provides foundations for statistical genetic data analysis, including a survey of fundamental concepts, primers on statistics and human evolution, and an introduction to polygenic scores. It then covers the practicalities of working with genetic data, discussing such topics as analytical challenges and data management. Finally, the book presents applications and advanced topics, including polygenic score and gene-environment interaction applications, Mendelian Randomization and instrumental variables, and ethical issues. The software and data used in the book are freely available and can be found on the book's website.

Machine Learning for Big Data Analysis CRC Press

A "genotype" is essentially an organism's full hereditary

information which is obtained from its parents. A "phenotype" is an organism's actual observed physical and behavioral properties. These may include traits such as morphology, size, height, eye color, metabolism, etc. One of the pressing challenges in computational and systems biology is genotype-to-phenotype prediction. This is challenging given the amount of data generated by modern Omics technologies. This “Big Data” is so large and complex that traditional data processing applications are not up to the task. Challenges arise in collection, analysis, mining, sharing, transfer, visualization, archiving, and integration of these data. In this Special Issue, there is a focus on the systems-level analysis of Omics data, recent developments in gene ontology annotation, and advances in biological pathways and network biology. The integration of Omics data with clinical and biomedical data using machine learning is explored. This Special Issue covers new methodologies in the context of gene–environment interactions, tissue-specific gene expression, and how external factors or host genetics impact the microbiome.

A Data Visualization Guide for Business Professionals Cambridge University Press

Recent advances in genomic sequencing technologies and big data analytics present a golden opportunity for bioinformatics, making it possible to efficiently analyze hundreds of thousands of individual genomes and identify statistically significant genetic determinants of disease. However, most research institutes lack sufficient genomic data to detect fine-grained signals crucial for understanding complex human diseases, and data sharing is often impractical due to strict privacy protections. By leveraging a cryptographic technique known as secure multiparty computation (MPC), researchers can securely cooperate on large-scale genomic studies without revealing sensitive data to any collaborators. In this thesis, we propose a public cloud-based computational framework that implements MPC for an essential genomic analysis workflow known as genome-wide association study (GWAS). By productionizing secure GWAS tools in an easy-to-use interface that abstracts away the technical challenges involved with implementing and running a protocol on several independent, geographically separated machines, we hope to enable researchers around the world to launch meaningful genomic studies with minimal overhead. We hope such efforts will prove instrumental towards the broader aim of establishing a single general-purpose platform for secure genomics research.

Contemporary Issues in Communication, Cloud and Big Data Analytics National Academies Press

Don't simply show your data—tell a story with it! Storytelling with Data teaches you the fundamentals of data visualization and how to communicate effectively with data. You'll discover the power of storytelling and the way to make data a pivotal point in your story. The lessons in this illuminative text are grounded in theory, but made accessible through numerous real-world examples—ready for immediate application to your next graph or presentation. Storytelling is not an inherent skill, especially when it comes to data visualization, and the tools at our disposal don't make it any easier. This book demonstrates how to go beyond conventional tools to reach the root of your data, and how to use your data to create an engaging, informative, compelling story. Specifically, you'll learn how to: Understand the importance of context and audience Determine the appropriate type of graph for your situation Recognize and eliminate the clutter clouding your information Direct your audience's attention to the most important parts of your data Think like a designer and utilize concepts of design in data visualization Leverage the power of storytelling to help your message resonate with your audience Together, the lessons in this book will help you turn your data into high impact visual stories that stick with your audience. Rid your world of ineffective graphs, one exploding 3D pie chart at a time. There is a story in your data—Storytelling with Data will give you the skills and power to tell it!

Proceedings of CSI 2015 Walter de Gruyter GmbH & Co KG

Leveraging Biomedical and Healthcare Data: Semantics, Analytics and Knowledge provides an overview of the approaches used in semantic systems biology, introduces novel areas of its application, and describes step-wise protocols for transforming heterogeneous data into useful knowledge that can influence healthcare and biomedical research. Given the astronomical increase in the number of published reports, papers, and datasets over the last few decades, the ability to curate this data has become a new field of biomedical and healthcare research. This book discusses big data text-based mining to better understand the molecular architecture of diseases and to guide health care decision. It will be a valuable resource for bioinformaticians and members of several areas of the biomedical field who are interested in understanding more about how to process and apply great amounts of data to improve their research. Includes at each section resource pages containing a list of available curated raw and processed data that can be used by researchers in the field Provides demonstrative and relevant examples that serve as a general tutorial Presents a list of algorithm names and computational tools available for basic and clinical researchers

4th International Conference, BDA 2015, Hyderabad, India, December 15-18, 2015, Proceedings IGI Global

Data in the genomics field is booming. In just a few years, organizations such as the National Institutes of Health (NIH) will host 50+ petabytes—or over 50 million gigabytes—of genomic data, and they’re turning to cloud infrastructure to make that data available to the research community. How do you adapt analysis tools and protocols to access and analyze that volume of data in the cloud? With this practical book, researchers will learn how to work with genomics algorithms using open source tools including the Genome Analysis Toolkit (GATK), Docker, WDL, and Terra. Geraldine Van der Auwera, longtime custodian of the GATK user community, and Brian O’Connor of the UC Santa Cruz Genomics Institute, guide you through the process. You’ll learn by working with real data and genomics algorithms from the field. This book covers: Essential genomics and computing technology background Basic cloud computing operations Getting started with GATK, plus three major GATK Best Practices pipelines Automating analysis with scripted workflows using WDL and Cromwell Scaling up workflow execution in the cloud, including parallelization and cost optimization Interactive analysis in the cloud using Jupyter notebooks Secure collaboration and computational reproducibility using Terra

High-Performance Big-Data Analytics Springer

The genomes in human body programs the blueprint of one’s life but the functions of those genomes nearly three billion genome bases are not known. The genome sequence in human being gives the fundamental rules for human biology. Science makes every effort to reveal the laws of nature and critical understanding of the biology. Scientists in the life-science field are seeking genetic variants associated with multifaceted set of observable characteristics to advance our understanding about genetics. Technological advancements are assisting the scientists to quickly create, store and analyze the data as fast as possible

and as efficient as possible. The NCBI and other organizations maintain genome sequences, proteins, RNA, DNA and other information of all species as well as their behavioral data. There is a lot and lot of data. Translating these data into useful insights which can be used for research and innovation is a main concern.

Proceeding of the International Conference on Computational and Bio Engineering, 2019, Volume 1 Springer Science & Business Media

Computational Genomics with R provides a starting point for beginners in genomic data analysis and also guides more advanced practitioners to sophisticated data analysis techniques in genomics. The book covers topics from R programming, to machine learning and statistics, to the latest genomic data analysis techniques. The text provides accessible information and explanations, always with the genomics context in the background. This also contains practical and well-documented examples in R so readers can analyze their data by simply reusing the code presented. As the field of computational genomics is interdisciplinary, it requires different starting points for people with different backgrounds. For example, a biologist might skip sections on basic genome biology and start with R programming, whereas a computer scientist might want to start with genome biology. After reading: You will have the basics of R and be able to dive right into specialized uses of R for computational genomics such as using Bioconductor packages. You will be familiar with statistics, supervised and unsupervised learning techniques that are important in data modeling, and exploratory analysis of high-dimensional data. You will understand genomic intervals and operations on them that are used for tasks such as aligned read counting and genomic feature annotation. You will know the basics of processing and quality checking high-throughput sequencing data. You will be able to do sequence analysis, such as calculating GC content for parts of a genome or finding transcription factor binding sites. You will know about visualization techniques used in genomics, such as heatmaps, meta-gene plots, and genomic track visualization. You will be familiar with analysis of different high-throughput sequencing data sets, such as RNA-seq, ChIP-seq, and BS-seq. You will know basic techniques for integrating and interpreting multi-omics datasets. Altuna Akalin is a group leader and head of the Bioinformatics and Omics Data Science Platform at the Berlin Institute of Medical Systems Biology, Max Delbrück Center, Berlin. He has been developing computational methods for analyzing and integrating large-scale genomics data sets since 2002. He has published an extensive body of work in this area. The framework for this book grew out of the yearly computational genomics courses he has been organizing and teaching since 2015.

Contemporary Issues in Communication, Cloud and Big Data Analytics Frontiers Media SA

Geneticists and molecular biologists have been interested in quantifying genes and their products for many years and for various reasons (Bishop, 1974). Early molecular methods were based on molecular hybridization, and were devised shortly after Marmur and Doty (1961) first showed that denaturation of the double helix could be reversed - that the process of molecular reassociation was exquisitely sequence dependent. Gillespie and Spiegelman (1965) developed a way of using the method to titrate the number of copies of a probe within a target sequence in which the target sequence was fixed to a membrane support prior to hybridization with the probe - typically a RNA. Thus, this was a precursor to many of the methods still in use, and indeed under development, today. Early examples of the application of these methods included the measurement of the copy numbers in gene families such as the ribosomal genes and the immunoglo bulin family. Amplification of genes in tumors and in response to drug treatment was discovered by this method. In the same period, methods were invented for estimating gene num bers based on the kinetics of the reassociation process - the so-called Cot analysis. This method, which exploits the dependence of the rate of reassociation on the concentration of the two strands, revealed the presence of repeated sequences in the DNA of higher eukaryotes (Britten and Kohne, 1968). An adaptation to RNA, Rot analysis (Melli and Bishop, 1969), was used to measure the abundance of RNAs in a mixed population.

Computing Systems and Approaches National Academies Press

Machine learning techniques are increasingly being used to address problems in computational biology and bioinformatics. Novel machine learning computational techniques to analyze high throughput data in the form of sequences, gene and protein expressions, pathways, and images are becoming vital for understanding diseases and future drug discovery. Machine learning techniques such as Markov models, support vector machines, neural networks, and graphical models have been successful in analyzing life science data because of their capabilities in handling randomness and uncertainty of data noise and in generalization. Machine Learning in Bioinformatics compiles recent approaches in machine learning methods and their applications in addressing contemporary problems in bioinformatics approximating classification and prediction of disease, feature selection, dimensionality reduction, gene selection and classification of microarray data and many more.

Knowledge Modelling and Big Data Analytics in Healthcare CRC Press

This book aims to help plant breeders by reviewing past achievements, currently successful practices, and emerging methods and techniques. Theoretical considerations are also presented to strike the right balance between being as simple as possible but as complex as necessary. The United Nations predicts that the global human population will continue rising to 9.0 billion by 2050. World food production will need to increase between 70-100 per cent in just 40 years. First generation bio-fuels are also using crops and cropland to produce energy rather than food. In addition, land area used for agriculture may remain static or even decrease as a result of degradation and climate change, despite more land being theoretically available, unless crops can be bred which tolerate associated abiotic stresses. Lastly, it is unlikely that steps can be taken to mitigate all of the climate change predicted to occur by 2050, and beyond, and hence adaptation of farming systems and crop production will be required to reduce predicted negative effects on yields that will occur without crop adaptation. Substantial progress will therefore be required in bridging the yield gap between what is currently achieved per unit of land and what should be possible in future, with the best farming methods and best storage and transportation of food, given the availability of suitably adapted cultivars, including adaptation to climate change. My book is divided into four parts: Part I is an historical introduction; Part II deals with the origin of genetic variation by mutation and recombination of DNA; Part III explains how the mating system of a crop species determines the genetic structure of its landraces; Part IV considers the three complementary options for future progress: use of sexual reproduction in further conventional breeding, base broadening and introgression; mutation breeding;

and genetically modified crops.

Related with Big Data Analytics In Genomics Springer:
• Temperature Conversion Worksheet Answers : click here