1. Record Nr.   UNINA9910337570903321

   Autore   Mangul Serghei

   Titolo   Hidden Treasures in Contemporary RNA Sequencing / / by Serghei
                       Mangul, Harry Taegyun Yang, Eleazar Eskin, Noah Zaitlen

   Pubbl/distr/stampa Cham : , : Springer International Publishing : , : Imprint : Springer, ,
                       2019

   ISBN   3-030-13973-5

   Edizione   [1st ed. 2019.]

   Descrizione fisica 1 online resource (99 pages)

   Collana   SpringerBriefs in Computer Science, , 2191-5768

   Disciplina   547.79
                       572.8633

   Soggetti   Bioinformatics
                       Genetics
                       Computational Biology/Bioinformatics
                       Genetics and Genomics

   Lingua di pubblicazione   Inglese

   Formato   Materiale a stampa

   Livello bibliografico   Monografia

   Nota di contenuto   Abstract -- Main Text -- Acknowledgement -- Disclosure --
                       References.

   Sommario/riassunto Advances in RNA-sequencing (RNA-seq) technologies have provided an
                       unprecedented opportunity to explore the gene expression landscape
                       across individuals, tissues, and environments by efficiently profiling the
                       RNA sequences present in the samples. When a reference genome
                       sequence or a transcriptome of the sample is available, mapping-based
                       RNA-seq analysis protocols align the RNA-seq reads to the reference
                       sequences, identify novel transcripts, and quantify the abundance of
                       expressed transcripts. The reads that fail to map to the human
                       reference, known as unmapped reads, are a large and often overlooked
                       output of standard RNA-seq analyses. Even in carefully executed
                       experiments, the unmapped reads can comprise a considerable fraction
                       of the complete set of reads produced, and can arise due to technical
                       sequencing produced by low-quality and error-prone copies of the
                       nascent RNA sequence being sampled. Reads can also remain
                       unmapped due to unknown transcripts, recombined B and T cell
                       receptor sequences, A-to-G mismatches from A-to-I RNA editing,

trans-splicing, gene fusion, circular RNAs, and the presence of non-host RNA sequences (e.g. bacterial, fungal, and viral organisms). Unmapped reads represent a rich resource for the study of B and T cell receptor repertoires and the human microbiome system—without incurring the expense of additional targeted sequencing. This book introduces and describes the Read Origin Protocol (ROP), a tool that identifies the origin of both mapped and unmapped reads. The protocol first identifies human reads using a standard high-throughput algorithm to map them onto a reference genome and transcriptome. After alignment, reads are grouped into genomic (e.g. CDS, UTRs, introns) and repetitive (e.g. SINEs, LINEs, LTRs) categories. The rest of the ROP protocol characterizes the remaining unmapped reads, which failed to map to the human reference sequences.