Vol. 25 no. 4 2009, pages 429
doi:10.1093/bioinformatics/btp037

BIOINFORMATICS FOR NEXT GENERATION SEQUENCING

The transforming aspect of the Human Genome Project was not the completion of the genome sequence itself, but rather the technologies that enabled, and were enabled by, the sequencing of that first reference genome. The evolution of ‘omic science through microarray transcriptomics, metabolomics, proteomics, and whole-genome SNP-omics has in many ways come full circle with a new focus on genomics and genome sequencing. Next-generation sequencing technologies have begun to revolutionise genomics and their effects are becoming increasingly widespread.

The 1000 genomes project (http://www.1000genomes.org/) will create a new map of genetic variation for our genome going far beyond the detail captured in the HapMap. Other projects are helping to catalogue genes involved in cancer, alternative splicing in different tissues and transcription factor binding, for example. The growing number of robust applications and the steadily falling cost for generating sequence-based data suggest that these next-generation technologies will continue to rapidly open new applications in the biological sciences and generate new opportunities for software and algorithm development. Given the vast amount of data produced (currently greater than a gigabase per run, with this constantly increasing as well), developing a sound data storage and management solution and creating informatics tools to effectively analyze the data are essential to successful application of the technology. During the past year, a large number of new software applications and algorithms have been developed to deal with this new data. A recent advert in Nature from the Illumina, one of the providers of next-generation sequencing technology, highlighted significant papers in the area of bioinformatics; our journal published 7 of the 16 listed papers. In addition to those cited in the advertisement, there have been many other tools and algorithms published in *Bioinformatics* that are relevant to next-generation sequencing applications. To celebrate this contribution we have gathered these together in a ‘Bioinformatics for Next Generation Sequencing’ virtual issue (http://bioinformatics.oxfordjournals.org/our_journals/bioinformatics/NextGenerationSequencing.html). This will be a living resource that we will continually update to include the very latest papers in this area to help researchers keep abreast of the latest developments.

To date, the majority of the papers have described methods to take the short sequences produced by the Illumina Genome Analyzer and Applied Biosystems SOLiD machines and align them to a reference genome. This is a crucial and basic requirement for many applications and a variety of techniques have been applied to make the tools sufficiently fast to deal with millions of sequences. We have also included papers that address the issue of assembly of these short reads. Now that there are many of these tools available the Bioinformatics community has begun to make applications that are useful for specific applications such as identifying likely sites of interaction in CHIP-seq. A summary of the inaugural collection is included in the Table 1. We sincerely hope that you find this resource useful and that the collected references lead to additional development in an area that we view as critical to the continued development of genomics and bioinformatics.

Alex Bateman and John Quackenbush

Table 1. Tools recently published in the journal

Author	Category	Title	Reference
De Bona *et al.*	Alignment	Optimal spliced alignments of short sequence reads	24:i174–i180
Prüfer *et al.*	Alignment	PatMaN: rapid alignment of short sequences to large databases	24(13):1530–1531
Jiang *et al*.	Alignment	SeqMap: mapping massive amount of oligonucleotides to the genome	24:2395–2396
Lin *et al.*	Alignment	ZOOM! Zillions of oligos mapped	24:2431–2437
Ondov *et al*.	Alignment	Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications	24(23):2776–2777
Malhis *et al.*	Alignment	Slider - Maximum use of probability information for alignment of short sequence reads and SNP detection	25(1):6–13
Li *et al.*	Alignment	SOAP: short oligonucleotide alignment program	24(5):713–714
Dalevi *et al*.	Alignment	Annotation of metagenome short reads using proxygenes	24(16):i7–i13
Hajirasouliha *et al.*	Alignment	Optimal pooling for genome re-sequencing with ultra-high-throughput short-read technologies	24(13):i32–i40
Miller *et al.*	Assembly	Aggressive Assembly of Pyrosequencing Reads with Mates	24(24):2818–2824
Zimin *et al.*	Assembly	Assembly reconciliation	24(1):42–45
Denisov *et al.*	Assembly	Consensus generation and variant detection by Celera Assembler	24(8):1035–1040
Warren *et al.*	Assembly	Assembling millions of short DNA sequences using SSAKE	23(4):500–501
Jeck *et al.*	Assembly	Extending assembly of short DNA sequences to handle error	23(21):2942–2944
Fejes *et al.*	CHIP-seq	FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology	24(15):1729–1730
Boyle *et al*.	CHIP-seq	F-Seq: a feature density estimator for high-throughput sequence tags	24(21):2537–2538