Katsuoka, Tohoku University, Japan; Masayuki Yamamoto, Tohoku University, Japan

Short Abstract:
Tohoku University Tohoku Medical Megabank Organization (ToMMo; http://www.megabank.tohoku.ac.jp/) was founded to establish an advanced medical system to foster the reconstruction from the Great East Japan Earthquake. The organization will develop a biobank that combines medical and genome information during the process of rebuilding the community medical system and supporting health and welfare in the Tohoku area. A blueprint for Tohoku University Tohoku Medical Megabank Organization is a ten-year project including three main activities: a biobank combining medical and genome information; an online platform for the coordination of community medical information; and training program designed for a varieties of highly specialized professionals and experts such as researchers of bioinformatics and science communicators. The biobank to be developed will be utilized to analyze the local heredity information so that it can establish an advanced medical system based on genome information with cutting-edge information and communication technology. The first goal of ToMMo is to understand the detailed genetic population background in this area including rare variants. Thus, this project applies deep coverage whole genome sequencing of thousands people who joined to this prospective genome cohort project within years. This poster presents the oral presenter's organizing part "the data management and the bioinformatics analysis of massive amount of high throughput sequencing data", and the research position availabilities of this very exciting project as a graduate student or a research staff.

Top

Poster N100

A statistical variant calling approach using pedigree information

Kaname Kojima, Tohoku Medical Megabank Organization, Tohoku University, Japan
Naoki Nariai, Tohoku Medical Megabank Organization, Tohoku University, Japan; Masao Nagasaki, Integrative Genomics, Japan

Short Abstract:
Due to the progress of next-generation sequencing technologies, whole genome sequencing for each individual becomes possible in practical time and with reasonable cost for the identification of disease associated mutations. Also, individual genome data from case-control study or study with pedigree analysis contribute to the elucidation of unknown disease mechanisms. Since accurate variant detection is required for the analysis of these genomes in a reliable manner, the development of accurate variant callers is demanded. In most of the variant callers, variants such as single nucleotide polymorphisms, insertions, and deletions are detected at each position in a reference genome from the information of mapped sequence reads. Since there exist positions with insufficient coverage of reads due to bias in the library preparation and mapping failures at short tandem repeat polymorphic sites or variable number of tandem repeat sites, reliable variant detection is challenging at these sites. Hence, variant callers that are robust to those errors even with low coverage data are demanded. We propose a new variant calling approach that considers pedigree information. Unlike variant callers considering individuals independently, our approach can use sequence read information of each individual for variant calling on other individuals by connecting read generation models of individuals based on pedigree information. Therefore, accurate variant calling and genotyping is expected even in insufficient read coverage sites. In performance evaluation with the HapMap CEU parent-offspring trio sequencing data, our approach outperformed existing approaches in accuracy on the agreement with SNP array genotyping results.

Top

Poster N101

Counting RNAseq reads: which way is better?

Felipe Rodrigues da Silva, Embrapa Informática Agropecuária, Brazil
Felipe Da Silva, Brazil; Roberto Willians Noda, Embrapa Milho e Sorgo, Brazil; Adhemar Zerlotini Neto, Embrapa Informática Agropecuária, Brazil; Francisco Pereira Lobo, Embrapa Informática Agropecuária, Brazil; Newton Portilho Carneiro, Embrapa Milho e Sorgo, Brazil

Short Abstract:
RNAseq presented a revolution on mRNA expression analysis. Microarrays where considered doomed and early experimental validation of RNAseq analysis findings has largely endorsed the common sense view of this technology would become the de facto standard on gene expression analysis. Some data RNAseq, however, seems to be very sensitive to the method used on its analysis. In this work we show the variation of results we've found while working with ~1 billion Illumina reads from drought tolerant Sorghum bicolor genotype in the presence and absence of the stress and compared results found for key genes already characterized.

Top

Poster N102

In *silico* screening for Antimicrobial Resistance genes in NGS Sequenced Bacterial Strains

Christian Rausch, Royal DSM, Netherlands

Short Abstract:
Royal DSM is a global science-based company active in health, nutrition and materials and a prominent player in industrial biotechnology. We have developed a dual approach to detect Antimicrobial Resistance (ARes) genes in bacterial genome sequences. The first part of the method implements a homology search at the DNA level which relies on mapping of the sequencing reads to a collection of known ARes genes and is an update of a method that was previously described by Brennedsen et al. (2011) accounting for 4 fold longer reads (150 bp) which are typically today obtained by Next gen sequencing machines like the Illumina MiSeq. In the second part of the method, the collection of ARes genes is searched at the protein level in thede novo assembly of the genome using translated BLAST. Because de novo assemblies built from NGS data today typically are not closed but consist of many contigs (in the order of 100 in our cases), the combination of both approaches allows to also assess genes that might be absent in the draft de novo assembly and takes advantage of increased sensitivity of homology searches at the protein instead of DNA level. We have applied our method to Lactobacillus bulgaricus strains, results will be given on the poster.

Top

Poster N103

Assembling the 20 Gb White Spruce Genome

Shaun Jackman, British Columbia Cancer Agency, Canada
Inanc Birol, Canada; Anthony Raymond, British Columbia Cancer Agency, Canada; Shaun D Jackman, British Columbia