Special Topics on Bioinformatics:
Intro to Biomolecular Structures and Genetics

Paper to discuss

L.Lehti

Probabilistic analysis of probe reliability in differential gene expression studies with short oligonucleotide arrays

Paper to report

Paper 7

Support vector machine classification and validation of cancer tissue samples using microarray expression data

First Paper

Reading Flow

1. Abstract
2. Introduction
3. Conclusion/Discussion: highlighting keywords
4. Methods

Analysis Flow:

1. Language/terminology understanding: Dictionary /sources
2. Keywords: highlight

First paper Overview
Reading Flow

1. Keywords

Probabilistic analysis of probe reliability in differential gene expression studies with short oligonucleotide arrays

Abstract—Probe defects are a major source of noise in gene expression studies. While existing approaches detect noisy probes based on external information such as genomic alignments, we introduce and validate a targeted probabilistic method for analyzing probe reliability directly from expression data and independently of the noise source. This provides insights into the various sources of probe-level noise and gives tools to guide probe design.

Index Terms—Applications, Biology and genetics, Parameter learning, Probabilistic algorithms

Probe ??

CNV-seq, a new method to detect copy number variation using high-throughput sequencing

Specificity ??

Sensitivity ??

Abstract

Background: DNA copy number variation (CNV) has been recognized as an important source of genetic variation. Array comparative genomic hybridization (aCGH) is commonly used for CNV detection, but the microarray platform has a number of inherent limitations.

Results: Here, we describe a method to detect copy number variation using shotgun sequencing, CNV-seq. The method is based on a robust statistical model that describes the complete analysis procedure and allows the computation of essential confidence values for detection of CNV. Our results show that the number of reads, not the length of the reads is the key factor determining the resolution of detection. This favors the next-generation sequencing methods that rapidly produce large amount of short reads.

Conclusion: Simulation of various sequencing methods with coverage between 0.1× to 8× show overall specificity between 91.7 – 99.9% and sensitivity between 72.2 – 96.5%. We also show the results for assessment of CNV between two individual human genomes.

First paper Overview
Reading Flow

Language understanding

Abstract

Specificity ??
Sensitivity ??

Probe ???

Wikipedia:

Hybridization probe

"*Fragment of DNA or RNA of variable length which is used in DNA or RNA samples to detect the presence of nucleotide sequences (the DNA target) that are complementary to the sequence in the probe*"

"*Statistical measures of the performance of a binary classification test.*

Sensitivity *measures the proportion of actual positives which are correctly identified as such*

Specificity *measures the proportion of negatives which are correctly identified*

A theoretical, optimal prediction can achieve 100% sensitivity

First paper Overview
Reading Flow

Introduction: We do get

- Background, current scientific situation or context
- What/where the problem is: previous reported works (citations from 7-12)
- What are the current methods to address the problem: Constrains
- Potential solutions by the authors:
  - Novelty
  - Validation
  - Results
  - Provided package: Robustness and Reproducibility !!!

Discussion/Conclusion: we do have

Summary of the introduction

Method overview:

Closing with the aim of the research: closing the circle

Directed to whom and Where can be applied