Number Ninjas

Bofan Chen, Anton Hung and Kevin Rouse

EXECUTIVE SUMMARY

Overview

Genetic sequencing is a next generation technology that enables scientists to determine an individual's susceptibility to certain diseases. It is the comparison of the expression of genes between healthy and affected patients.

The Problem

Given two clusters of samples, healthy vs. affected, we needed to identify which genes were differentially expressed and driving the clustering pattern. The challenge was to determine which cluster of samples is healthy and which is sick, and what ailment they are suffering from.

The Solution

Step 1: Wilcoxon signed-rank test with Bonferroni multiple test correction was used to determine our 30 differentially expressed genes

Step 2: Translating our Ensembl gene IDs to gene names

Step 3: compare up and down-regulation of gene expression in our groups to trends found in the literature to identify a. What ailment are the subjects suffering from? and b. Which group is the affected group?

Highlights

We identified 30 differentially expressed genes between the healthy and sick groups. We identified the disease to be colorectal cancer and used literature information on what genes are up/down-regulated in colorectal cancer to determine which group was healthy vs sick.

Application to healthcare

When given a patient's genetic sequence, we can run a K-nearest neighbours machine learning algorithm. This employs euclidean distance scoring as well as feature importance to classify the patient to the nearest cluster, whether they are healthy or sick, or at risk of developing the disease in the future.
If they are classified to the affected group, then the physician can continue to follow up on pay attention to any onset of symptoms related to colorectal cancer.

Advancement in diagnostic technologies in clinical practice are extremely valuable because they help us to identify potentially fatal diseases very early on in their development, enabling physicians to treat them before they have spread and become irreversible.
Gene sequencing could become a part of your routine doctor's visit. evaluation of your genetic sequence combined with comparison against a human reference genome reveals differentially expressed genes in your genetic makeup. This can bring a doctor's attention to any potential warning signs of future health complications