

8/10/2012

Finding biomarkers for lung cancer using NGS analysis of RNA sequences



Soumya Nayak
Mentor: Dr S. Srinivasan

Contents

- **The Problem**
- **Our Approach to the solution**
- **Nature of the Dataset**
- **Computing Read per Gene**
- **Computing RPKM**
- **Results**
- **Microarray Analysis**
- **The NGS Engine: A tool**

The Problem

- Find Biomarkers for lung cancer
- Why is it interesting?
 - Prediction of cancer
 - Understand gene expression in cancer
 - Explain the effect of smoking
 - Data, data and more data!

Our Approach to Solution

**Download
Datasets**

**Aligning
it to an
index
genome**

**Storing
the
aligned
data to
MySQL
server**

**Compute
Reads per
gene per
each
sample
using an
annotation
file**

**Compute
(RPKM
Value)**

**Stat-
analysis
and
results**

Nature of the dataset

- Downloaded from SRA (SRP006676).
- Data is taken from 4 different types of people
 - Healthy non smokers
 - Current Smokers
 - Smokers with cancer
 - Smokers without cancer
- Two different approaches were used on same sample to collect the dataset
 - NuGen
 - Illumina

Aligning each sample to reference genome

- Reference genome used -Hg19
- Tool used-Bowtie
- Parameters used for bowtie
 - best,-v,-m
- An example of map file:

```
SRR192335.7 HWUSI-EAS1671_0001:7:1:1030:18590 length=75 - chr11
76262558
```

```
CTATAGTTAATANNTCCCACTAGAGTGACACTGAAGATTAAACACAAGCATTC
ATAAGATGCGCTGATCTCT
```

```
G#####@B;BABBBBBA>C@BBBBC>BAB@CACCBBCBCCACCCB<CC
BCCCCCBBCCC@CCC
```

```
C=C 0 61:A>C,62:A>G
```

Data entry to MySQL

23 tables corresponding to each chromosome is created in MySQL database.

An Example in table chr1:

sample	read_id	strand	start	end	seq
HNS1	SRR19233 3.47	+	207103 170	20710324 4	NTACCTGTCACAG CATCTAGCAAAG CAC

Calculating Reads per Gene

- What is it?
- How to calculate?
 - Bowtie output data is compared with an annotation file
 - The final output is as follows:

Acc_id	chr	HNS1	HS1	SWOC1	SWC1	HNS2	HS2	SWOC2	SWC2	GENE_NAME
NR_015368	CHR1	0	0	0	0	0	0	0	0	NULL
uc001aar.1	CHR1	123	256	98	16	23	12	34	6	LOC6438

Calculating RPKM

- RPKM is (Reads per Kilobase of exon model per Million mapped reads)

- Why RPKM?

- Formula:

RPKM =

$$\frac{\text{Total exon reads}}{\text{Mapped reads millions} \times \text{Exon-length Kilobases}_e}$$

Results

- Comparison between two methods

Many non-coding RNAs were captured in Nugen Protocol, i.e fold change was negative

- Comparison between different samples

GENE_NAME	HNS1-HS1	HNS1-SWC1	median	"HS1-SWC1
ZFY	-8.47	-8.54	-8.5	-0.07
CYorf15B	-8.22	-9.07	-8.64	-0.85
BCAN	6.17	6.8	6.48	0.71

Pathway Analysis

- **NEUROACTIVE LIGAND RECEPTOR INTERACTION**
- **ANACHIDONIC ACID METABOLISM**
- **SYSTEMATIC LUPUS ERYTHEMATOS**
- **MELANOGENESIS**
- **CALCIUM SIGNALLING PATHWAY**

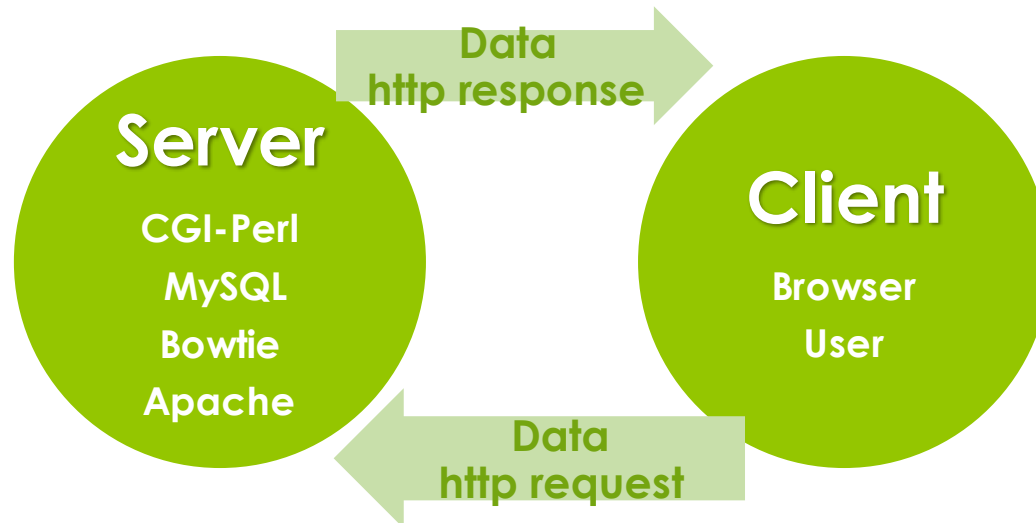
Microarray Analysis of Prostate Cancer

- Pipeline used
- Dataset taken : GSE6919
- The experiment is about: Expression Data from Normal and Metastatic Prostate Tumor Tissues
- For metastatic tumor is categorized into different types:
 - prostate tumor metastases in liver
 - prostate tumor metastases in para tracheal lymph node
 - prostate tumor metastases in para aortic lymph node
 - prostate tumor metastases in retro peritoneal lymph node

The output of the analysis

The NGS Engine: A tool

- A Web-based tool for NGS
- Automates the analysis of NGS gene expression process
- Client-Server Architecture



Welcome to The NGS Engine

References

NGS

- <http://bowtie-bio.sourceforge.net/index.shtml>
- http://www.ncbi.nlm.nih.gov/pubmed?LinkName=sra_pubmed&from_uid=69947
- [Beane J](#), et al, -2011 Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq [Cancer Prev Res \(Phila\)](#).
<http://dev.mysql.com/doc/refman/5.0/en/tutorial.html>
- <http://perl-tutorial.org/>
- http://www.clcbio.com/manual/genomics/Definition_RPKM.html

DATA

- <http://www.ncbi.nlm.nih.gov/sra/>

TOOL

- <http://perl-tutorial.org/>
- <http://www.w3schools.com/html/default.asp>
- <http://www.w3schools.com/css/>
- <http://www.milonic.com/apachesetup.php>
- <http://tldp.org/LDP/LG/issue12/server.html>
- <http://stackoverflow.com/questions/6852376/architecture-for-a-local-html-gui-with-a-perl-back-end>