# 第5回バイオインフォマティクス実習コース

横浜市大　先端医科学研究センター
バイオインフォマティクス研究室
室長　　田村智彦
准教授　　中林潤
免疫学　　藩龍馬

- RNA-seqデータ解析

# RNA-seqデータ解析の手順

- シーケンス
- ゲノム上にマッピング
- 発現量に換算
- 発現解析

# マッピングツール
# tophat

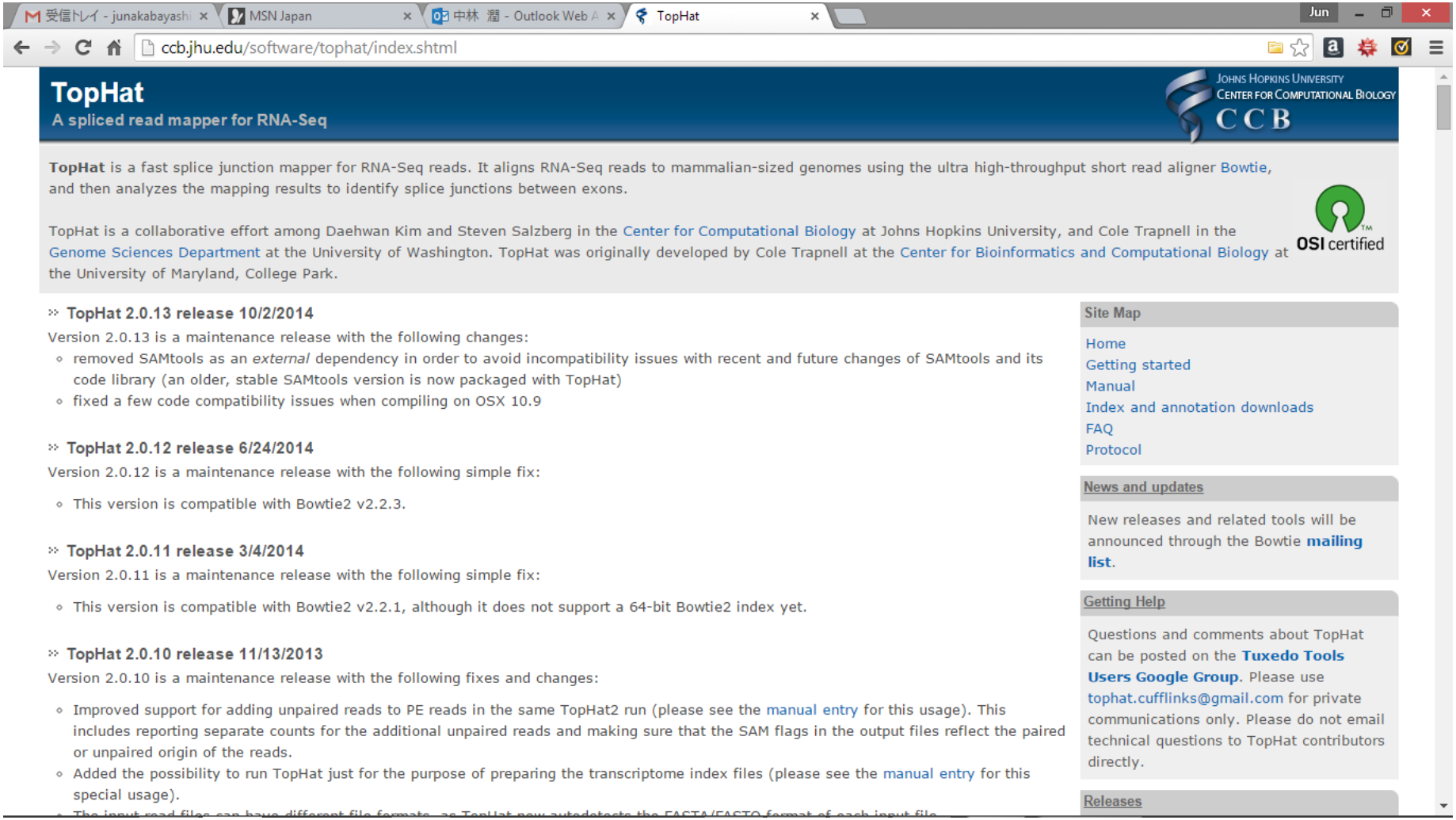
- Johns Hopkins University
  Center for Computational Biology

- http://ccb.jhu.edu/software/tophat/index.shtml

- Transcriptome解析用マッピングツール
  Bowtie2を呼び出してマッピング
  スプライスジャンクションを予測する

# TopHat

http://ccb.jhu.edu/software/tophat/index.shtml



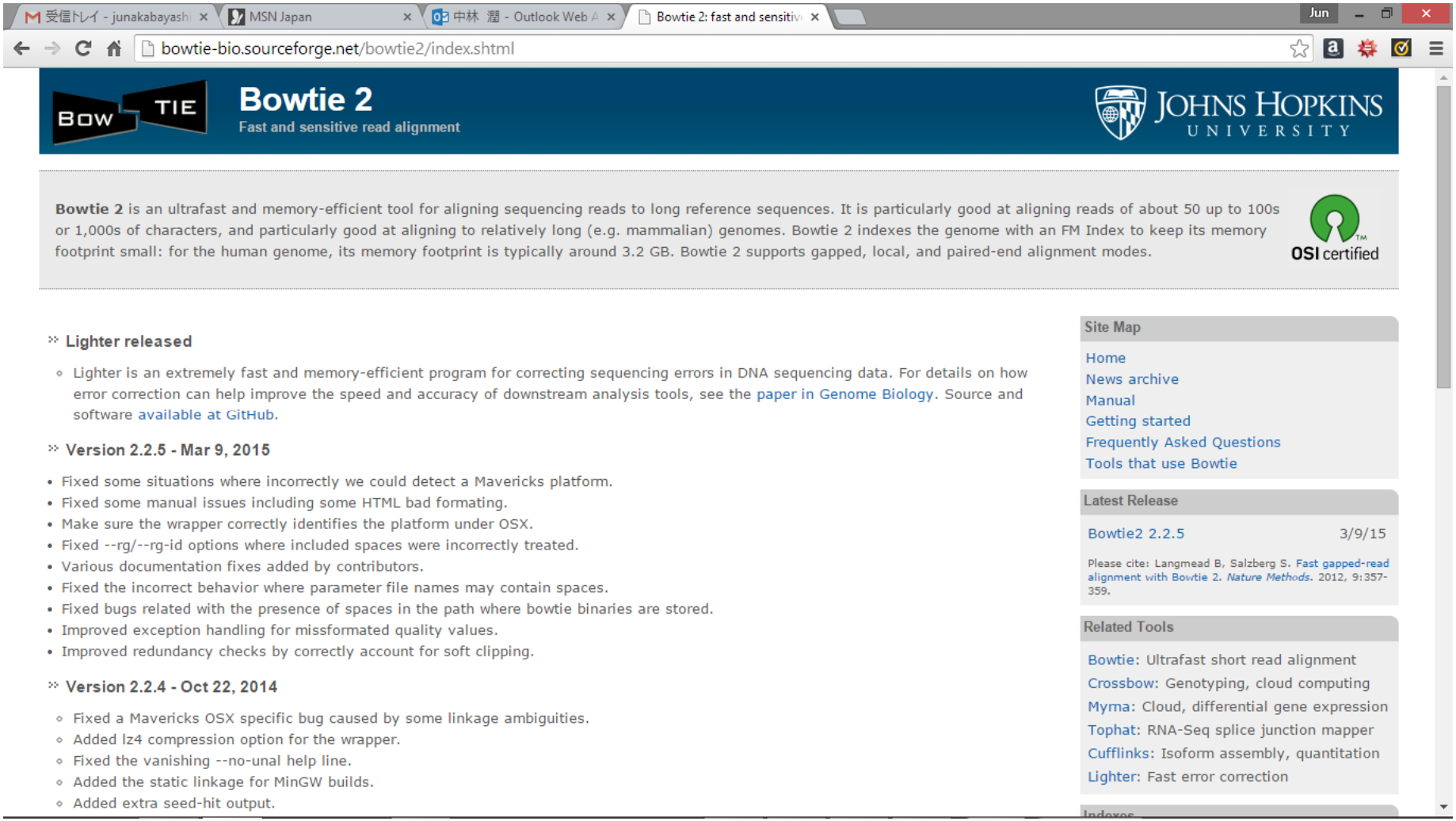
# マッピングツール
# Bowtie2

- John Hopkins University

- http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

# Bowtie2

http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

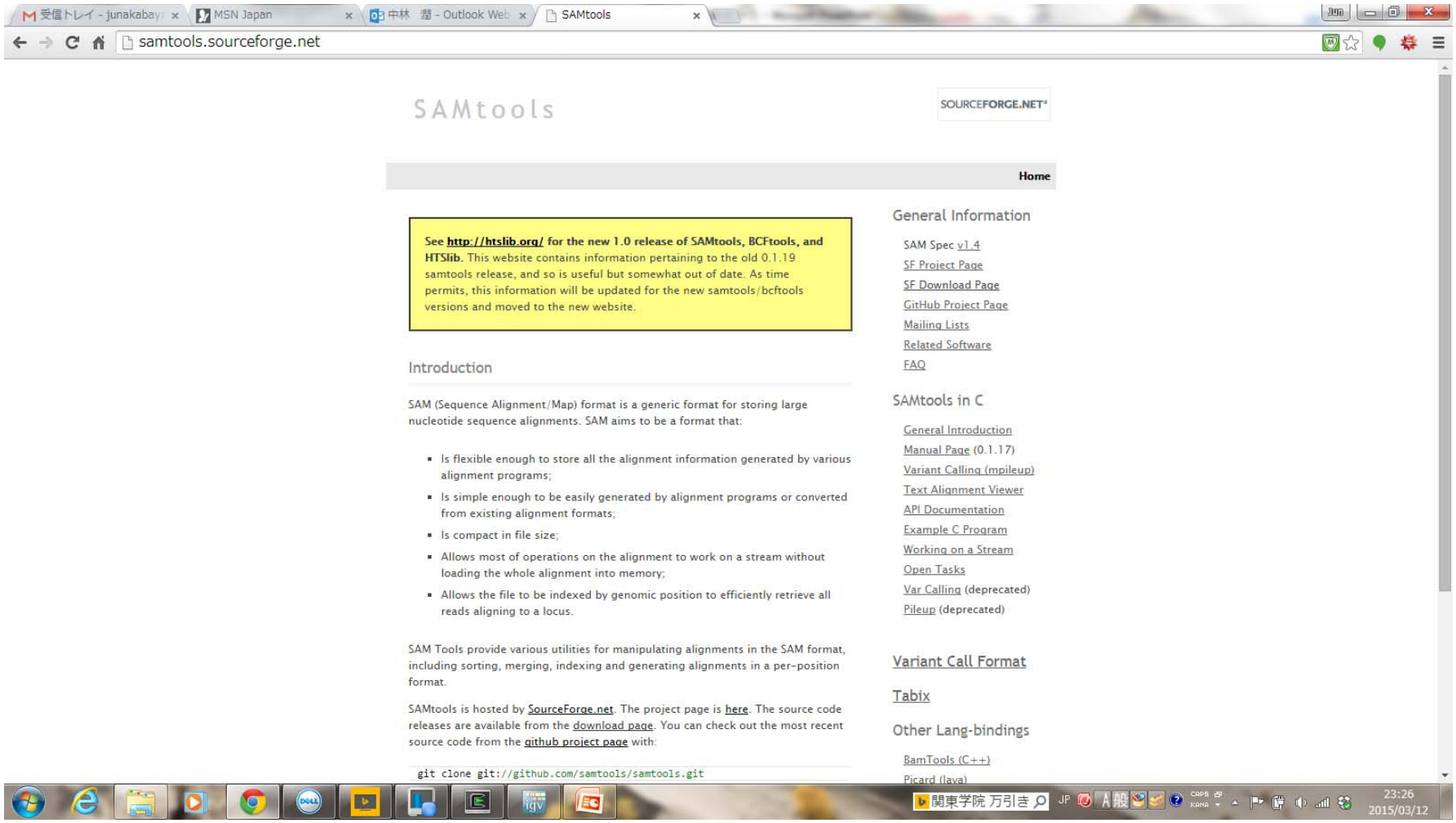


# samtools

- http://samtools.sourceforge.net/
- sam→bam変換
- sam fileのsort
- index作成

# SAMtools

http://samtools.sourceforge.net/



# Integrative Genomics Viewer

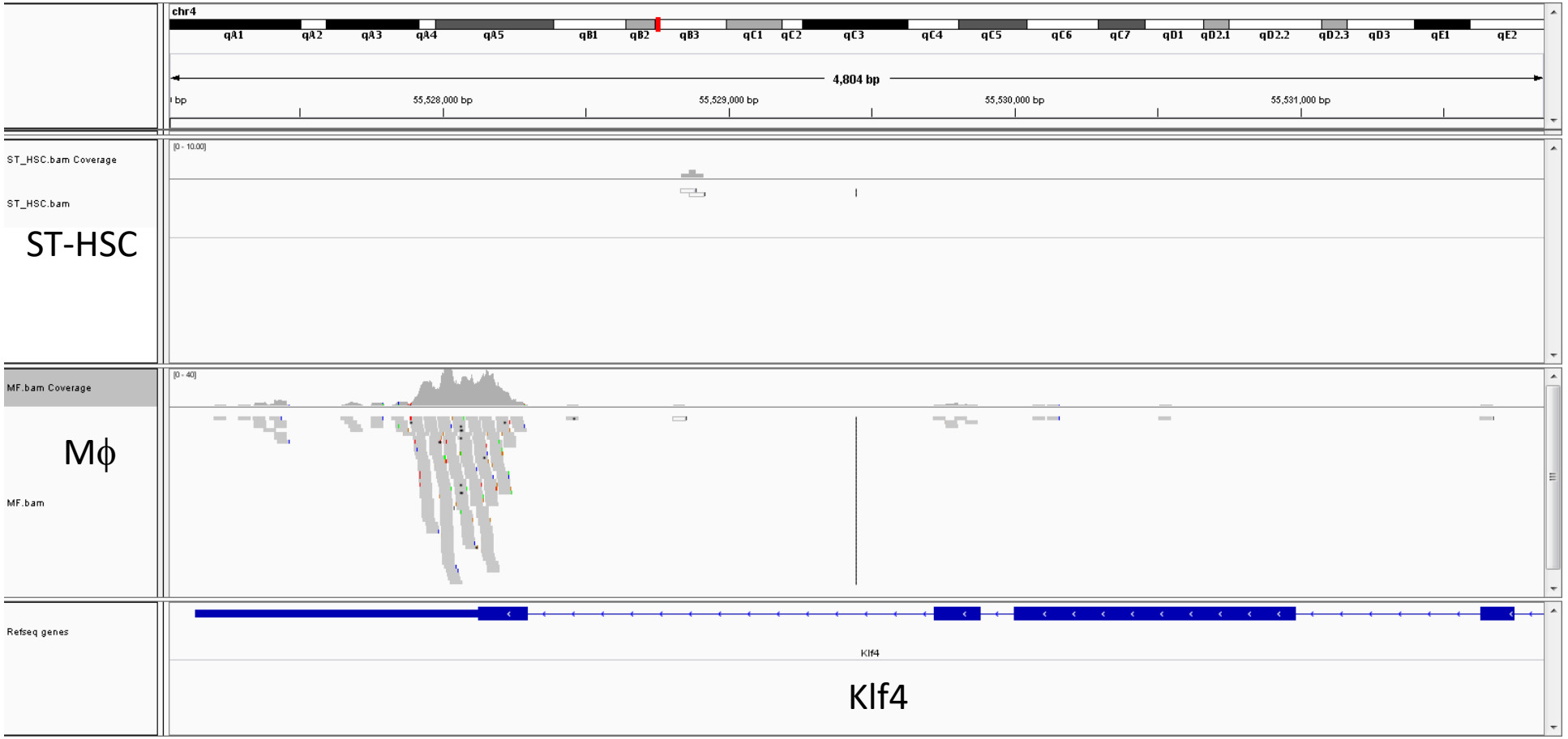
Broad institute
http://broadinstitute.org/igv/

GSE60101から
ST_HSC、Mϕの
遺伝子発現プロファイル
FASTQ fileを取得

tophatで
マッピング

bam file
samtoolsで
index作成

integrative genomics viewerで
表示



# cufflinks
# 発現定量

- マッピングデータを発現量に換算

⬇

RPKM
reads per kilobase of exon
per million mapped sequence reads
マッピングされたリード数をエクソン長と総リード数で正規化した値

$$RPKM = \frac{X_t}{l_t N} \times 10^9$$

$X_t$ : 転写物tにマップされたリード数
$l_t$ : 転写物tの長さ
$N$ : 総リード数

# R package “cummeRbund”

R console

```
> Sys.setenv(http_proxy = “http://proxy.yokohama-cu.ac.jp:8080”)
> source(“http://bioconductor.org/biocLite.R”)
> biocLite(“cummeRbund”)
> library(cummeRbund)
> x <- readCufflinks()
```

- proxyの設定
- biocLite.Rの設定
- パッケージ“cummeRbund”の読み込み
- 変数xに発現量データ（cuffdiffの出力）を格納

# R package “cummeRbund”

R console

```
> y <- genes(x)
> csDensity(y)
> csScatter(y, “q1”, “q2”)
> csBoxplot(y)
> csDendro(y)
```

- 遺伝子ごとの発現量を取得し、変数yに格納
- density plot, dendrogram, scatter plot, boxplot, dendrgramを作図

# R package “cummeRbund”

R console

```
> z <- fpkmMatrix(y)
> write.table(z, “FPKM_GSE60101.txt”, quote=F, sep=“¥t”)
```

- 発現量データを変数zに格納
- タブ区切りテキストファイルとして出力

アンケートにご協力をお願いいたします。