

# 【目 次】

- **遺伝子発現解析(1)**

- 各種ソフトウェアの説明とインストール

- **遺伝子発現解析(2)**

- レファレンス（参照）配列、解析用SRAファイルの入手とファイルフォーマットの理解
- シーケンスデータのクオリティチェックとトリミング

- **遺伝子発現解析(3)**

- RNAスプライシングアライメント  
(TopHat)
- マッピングデータのフォーマットの理解

# インストールするソフトウェア

- R
- RStudio

## **ターミナルを使って**

- コマンドライン・デベロッパ・ツール
- Homebrew
- wget
- bowtie/Tophat
- Cufflinks

## **Rを使って**

- cummeRbund

# wgetのインストール

指定したファイルをダウンロードするソフトウェア。  
ファイルをダウンロード中に中断してしまった場合でも再開できる機能が付いている。

```
$ brew install wget
```

wgetがインストールされたか確認しよう

```
$ which wget
```

```
/usr/local/bin/wget
```

wgetの使い方は非常に簡単！

```
$ wget Filename
```

```
$ wget -h
```

# wgetのオプション

```
GNU Wget 1.16.3, a non-interactive network retriever.  
Usage: Wget [OPTION]... [URL]...  
  
Mandatory arguments to long options are mandatory for short options too.  
  
Startup:  
  -V, --version           display the version of Wget and exit  
  -h, --help              print this help  
  -b, --background        go to background after startup  
  -e, --execute=COMMAND   execute a '.wgetrc'-style command  
  
Logging and input file:  
  -o, --output-file=FILE  log messages to FILE  
  -a, --append-output=FILE append messages to FILE  
  -q, --quiet             quiet (no output)  
  -v, --verbose            be verbose (this is the default)  
  -nv, --no-verbose         turn off verboseness, without being quiet  
  --report-speed=TYPE      output bandwidth as TYPE. TYPE can be bits  
  -i, --input-file=FILE    download URLs found in local or external FILE  
  -F, --force-html          treat input file as HTML  
  -B, --base=URL            resolves HTML input-file links (-i -F)  
                            relative to URL  
  --config=FILE            specify config file to use  
  --no-config              do not read any config file  
  
Download:  
  -t, --tries=NUMBER       set number of retries to NUMBER (0 unlimits)  
                            --retryconnrefused  
  -O, --output-document=FILE write documents to FILE  
  -nc, --no-clobber        skip downloads that would download to  
                            existing files (overwriting them)  
  -c, --continue            resume getting a partially-downloaded file  
                            --start-pos=OFFSET  
                            --progress=TYPE  
                            --show-progress  
  -N, --timestamping        don't re-retrieve files unless newer than  
                            local
```

ダウンロードの再開は以下のコマンドができる

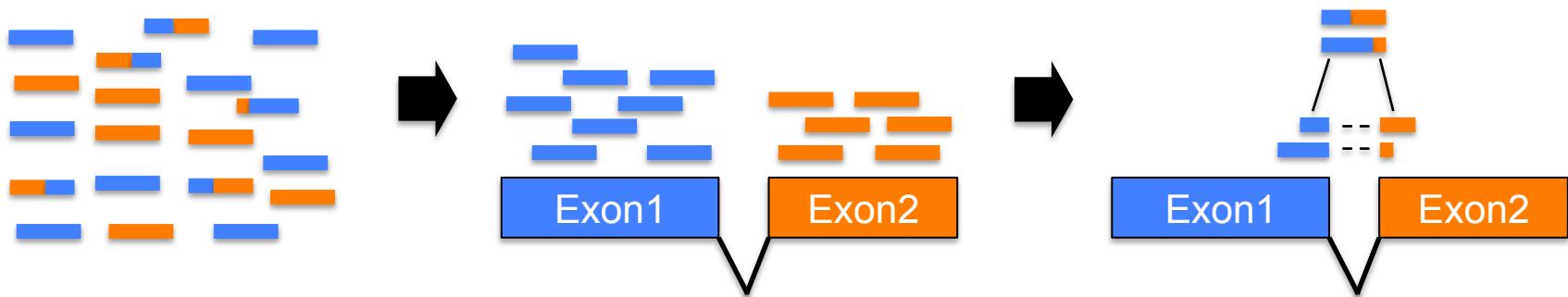
```
$ wget -c Filename
```

# bowtie/TopHat

遺伝子発現解析の際, 解析対象の配列をレファレンス配列にマッピングするための代表的なソフトウェアの1つ.

このソフトウェアの優れた点は, スプライスジャングクションを推定してくれることである.

RNA-Seq Data



# bowtie/TopHatのインストール

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

**TopHat**  
A spliced read mapper for RNA-Seq

JOHNS HOPKINS UNIVERSITY  
CENTER FOR COMPUTATIONAL BIOLOGY  
**CCB**

**TopHat** is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner **Bowtie**, and then analyzes the mapping results to identify splice junctions between exons.

TopHat is a collaborative effort among Daehwan Kim and Steven Salzberg in the [Center for Computational Biology](#) at Johns Hopkins University, and Cole Trapnell in the [Genome Sciences Department](#) at the University of Washington. TopHat was originally developed by Cole Trapnell at the [Center for Bioinformatics and Computational Biology](#) at the University of Maryland, College Park.

**OSI certified**

**TopHat 2.1.0 release 6/29/2015**

- TopHat-Fusion algorithm improvements for more sensitive and accurate discovery of fusions, thanks to contributions from Gordon Bean and Ryan Kelley at Illumina.
  - This release implements a new algorithm for counting fusion-supporting read pairs that reduces the number of false-positive potential fusions. This algorithm computes the inner distance between read pairs by first converting the pair positions to transcript coordinates using the transcript information in refGene.txt and ensGene.txt. Pairs with small inner distance (suggesting the pair could come from a plausible pair-end insert) are counted as supporting evidence for the fusion. The default threshold for the inner distance is 250 base pairs; this parameter can be set using the --fusion-pair-dist <int> flag.
- fixed a few issues with GFF parsing of some annotation files
- fixed a runtime-error when using --no-discordant option.

Several fixes/improvements thanks to contributors on GitHub:

- new --max-num-fusions option allowing the user to specify the maximum number of reported fusions in tophat-fusion-post
- adjusting lower limit for --fusion-multipairs
- fixed a few typos, cleaning up python code etc.

**TopHat source code moved to GitHub 3/31/2015**

TopHat is now available as a public GitHub repository where users are welcome to submit bug reports (issues) and developers are encouraged to submit patches (pull requests).

**TopHat 2.0.14 release 3/24/2015**

Version 2.0.14 is a maintenance release with the following changes:

- pipeline speed improvements thanks to contributions from Véronique Legrand and Michaël Pressigout of Institut Pasteur
- added support for xz compressed read files (thanks to a patch submitted by Ashton Trey Belew)
- applied a couple of Python fixes to prevent potential issues with package handling and some file operations
- fixed a potential linking issue where the wrong libbam.a library could have been linked when building from source

**TopHat 2.0.13 release 10/2/2014**

Version 2.0.13 is a maintenance release with the following changes:

- removed SAMtools as an *external* dependency in order to avoid incompatibility issues with recent and future changes of SAMtools and

**Site Map**

- [Home](#)
- [Getting started](#)
- [Manual](#)
- [Index and annotation downloads](#)
- [FAQ](#)
- [Protocol](#)

**News and updates**

New releases and related tools will be announced through the Bowtie [mailing list](#).

**Getting Help**

Questions and comments about TopHat can be posted on the [Tuxedo Tools Users Google Group](#). Please use [tophat.cufflinks@gmail.com](mailto:tophat.cufflinks@gmail.com) for private communications only. Please do not email technical questions to TopHat contributors directly.

**Releases**

version 2.1.0	6/29/2015
<a href="#">Source code</a>	
<a href="#">Linux x86_64 binary</a>	
<a href="#">Mac OS X x86_64 binary</a>	

# bowtie/TopHatのインストール

Homebrewでインストールしよう

98ページ参照

# \$ tophat

```
tophat:  
TopHat maps short sequences from spliced transcripts to whole genomes.  
  
Usage:  
    tophat [options] <bowtie_index> <reads1[,reads2,...]> [reads1[,reads2,...]] \  
                [quals1,[quals2,...]] [quals1[,quals2,...]]  
  
Options:  
-v/--version  
-o/--output-dir          <string>   [ default: ./tophat_out ]  
--bowtie1  
-N/--read-mismatches     <int>      [ default: 2 ]  
--read-gap-length         <int>      [ default: 2 ]  
--read-edit-dist          <int>      [ default: 2 ]  
--read-realign-edit-dist  <int>      [ default: "read-edit-dist" + 1 ]  
-a/--min-anchor           <int>      [ default: 8 ]  
-m/--splice-mismatches    <0-2>     [ default: 0 ]  
-i/--min-intron-length    <int>      [ default: 50 ]  
-I/--max-intron-length    <int>      [ default: 500000 ]  
-g/--max-multihits        <int>      [ default: 20 ]  
--suppress-hits  
-x/--transcriptome-max-hits <int>      [ default: 60 ]  
-M/--prefilter-multihits  
                            ( for -G/--GTF option, enable  
                            an initial bowtie search  
                            against the genome )  
--max-insertion-length    <int>      [ default: 3 ]  
--max-deletion-length     <int>      [ default: 3 ]  
--solexa-quals  
--solexa1.3-quals          (same as phred64-quals)  
--phred64-quals            (same as solexa1.3-quals)  
-Q/--quals  
--integer-quals  
-C/--color                 (Solid - color space)  
--color-out  
--library-type             <string>   (fr-unstranded, fr-firststrand,  
                                         fr-secondstrand)
```

# Cufflinksのインストール

TopHat によるマッピング結果をもとに, 遺伝子発現構造を予測し, 遺伝子発現の定量およびサンプル間の発現量の差を算出するためのソフトウェアである

<http://cole-trapnell-lab.github.io/cufflinks/>

INSTALL    MANUAL    GETTING STARTED    TOOLS    HELP    HOW IT WORKS    PROTOCOL    BENCHMARKS    CODE     FEED

## Cufflinks

*Transcriptome assembly and differential expression analysis for RNA-Seq.*

Cufflinks assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples. It accepts aligned RNA-Seq reads and assembles the alignments into a parsimonious set of transcripts. Cufflinks then estimates the relative abundances of these transcripts based on how many reads support each one, taking into account biases in library preparation protocols.

Cufflinks was originally developed as part of a collaborative effort between the [Laboratory for Mathematical and Computational Biology](#), led by Lior Pachter at UC Berkeley, Steven Salzberg's [computational genomics group](#) at the Institute of Genetic Medicine at Johns Hopkins University, and [Barbara Wold's lab](#) at Caltech. The project is now maintained by [Cole Trapnell's lab](#) at the University of Washington.

# Cufflinksのインストール

Homebrewでインストールしよう

98ページ参照

# Cufflinksのインストール

```
$ cufflinks
```

```
cufflinks v2.2.1
linked against Boost version 104900
-----
Usage: cufflinks [options] <hits.sam>
General Options:
-o/--output-dir           write all output files to this directory      [ default: ./ ]
-p/--num-threads          number of threads used during analysis      [ default: 1 ]
--seed                     value of random number generator seed      [ default: 0 ]
-G/--GTF                  quantitate against reference transcript annotations
-g/--GTF-guide            use reference transcript annotation to guide assembly
-M/--mask-file             ignore all alignment within transcripts in this file
-b/--frag-bias-correct    use bias correction - reference fasta required      [ default: NULL ]
-u/--multi-read-correct   use 'rescue method' for multi-reads (more accurate)  [ default: FALSE ]
--library-type              library prep used for input reads          [ default: below ]
--library-norm-method     Method used to normalize library sizes        [ default: below ]

Advanced Abundance Estimation Options:
-m/--frag-len-mean         average fragment length (unpaired reads only)  [ default: 200 ]
-s/--frag-len-std-dev      fragment length std deviation (unpaired reads only)  [ default: 80 ]
--max-mle-iterations       maximum iterations allowed for MLE calculation  [ default: 5000 ]
--compatible-hits-norm    count hits compatible with reference RNAs only  [ default: FALSE ]
--total-hits-norm          count all hits for normalization          [ default: TRUE ]
--num-frag-count-draws    Number of fragment generation samples        [ default: 100 ]
--num-frag-assign-draws   Number of fragment assignment samples per generation  [ default: 50 ]
--max-frag-multihits      Maximum number of alignments allowed per fragment  [ default: unlim ]
--no-effective-length-correction  No effective length correction  [ default: FALSE ]
--no-length-correction    No length correction          [ default: FALSE ]
-N/--upper-quartile-norm  Deprecated, use --library-norm-method      [ DEPRECATED ]
--raw-mapped-norm          Deprecated, use --library-norm-method      [ DEPRECATED ]
```

# cummeRbundのインストール



RStudio

※ 使用するのは6コマ目(明日) 99ページ参照

The screenshot shows the RStudio interface with three main panes:

- Editor (Left):** Displays the code area with the text "エディタ画面".
- Environment (Top Right):** Displays the global environment with the message "Environment is empty".
- Console (Bottom Left):** Displays the R console output with the message "オブジェクト一覧画面".

The console output includes:

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件下に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()'あるいは'licence()'と入力してください。

R は多くの貢献者による共同プロジェクトです。
詳しくは'contributors()'と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()'と入力してください。

'demo()'と入力すればデモをみることができます。
'help()'とすればオンラインヘルプが出ます。
'help.start()'で HTML ブラウザによるヘルプが開かれます。
'q()'と入力すれば R を終了します。
```

# スリープモードの解除

解析の途中でスリープモードになってしまふと  
止まってしまう。



100ページ参照

システム環境設定

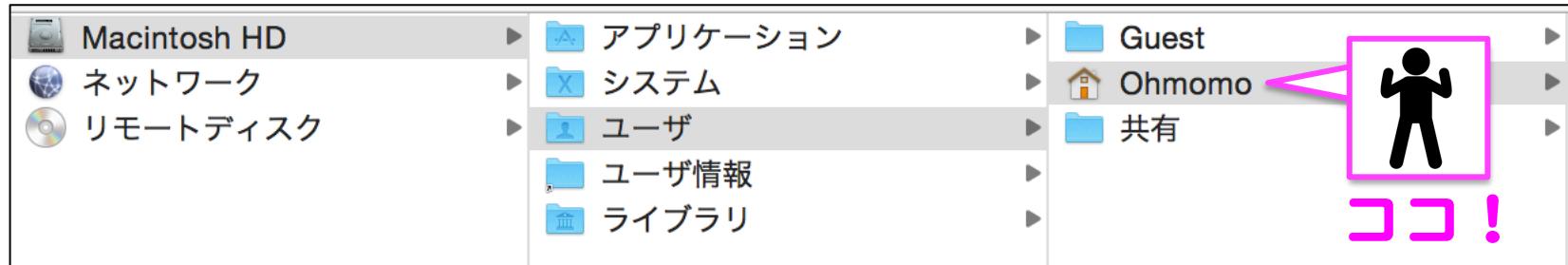
# 遺伝子発現解析(2)

- ・レファレンス（参照）配列、解析用SRAファイルの入手とファイルフォーマットの理解
- ・シークエンスデータのクオリティチェックとトリミング

# 前準備

home directoryへ移動

```
$ cd
```



解析データを入れるDirectoryを作成しよう

```
$ cd expression
```

# レファレンス配列データの入手

101ページ参照



## 1. Dry解析教本専用サーバからダウンロード

([http://shujunsha.com/NGS\\_DAT/Lv2\\_2/iGenome](http://shujunsha.com/NGS_DAT/Lv2_2/iGenome))



## 2. 公開データのダウンロード

### Ensembl

(<http://www.ensembl.org/index.html>)

#### <入手するファイル>

- Fastaファイル (genome.fa)
- GTFファイル (genes.gtf)



### iGenome

([https://support.illumina.com/sequencing/sequencing\\_software/igenome.html](https://support.illumina.com/sequencing/sequencing_software/igenome.html))

# 1. Dry解析教本専用サーバからダウンロード

## 現在、サーバの挙動がおかしくなっています...

[http://shujunsha.com/NGS\\_DAT/](http://shujunsha.com/NGS_DAT/)

### Index of /NGS\_DAT

Name	Last modified	Size	Description
<a href="#">Parent Directory</a>		-	
<a href="#">Lv1_2/</a>	08-Dec-2015 11:32	-	
<a href="#">Lv1_3/</a>	17-Jul-2015 14:49	-	
<a href="#">Lv2_2/</a>	28-Jul-2015 14:25	-	
<a href="#">Lv2_3/</a>	28-Jul-2015 15:01	-	
<a href="#">Lv2_5/</a>	07-Sep-2015 18:43	-	
<a href="#">Lv2_operation_manual/</a>	30-Sep-2015 17:12	-	
<a href="#">Lv3_B/</a>	04-Aug-2015 15:41	-	
<a href="#">Lv3_C/</a>	06-Aug-2015 12:35	-	
<a href="#">Lv3_D/</a>	06-Aug-2015 12:35	-	
<a href="#">Lv3_G/</a>	04-Aug-2015 16:50	-	
<a href="#">Lv3_H/</a>	30-Aug-2015 15:22	-	
<a href="#">Lv3_I/</a>	04-Aug-2015 16:50	-	
<a href="#">Lv3_J/</a>	04-Aug-2015 16:50	-	
<a href="#">Lv3_K/</a>	04-Aug-2015 15:42	-	
<a href="#">Lv3_L/</a>	04-Aug-2015 18:44	-	
<a href="#">Lv3_M/</a>	06-Aug-2015 12:36	-	
<a href="#">Lv3_N/</a>	06-Aug-2015 12:36	-	
<a href="#">Lv3_O/</a>	06-Aug-2015 12:36	-	
<a href="#">license</a>	16-Jul-2015 18:21	0	
<a href="#">sitemap</a>	16-Jul-2015 18:21	0	

### Index of /NGS\_DAT/Lv2\_2/iGenome

/Homo\_sapiens/NCBI/build37.2/Annotation  
/Archives/archive-2014-06-02-13-47-29/Genes

Name	Last modified	Size	Description
<a href="#">Parent Directory</a>		-	
<a href="#">genes.gtf</a>	03-Jun-2014 07:32	102M	

### Index of /NGS\_DAT/Lv2\_2/iGenome

/Homo\_sapiens/NCBI/build37.2/Sequence  
/Bowtie2Index

Name	Last modified	Size	Description
<a href="#">Parent Directory</a>		-	
<a href="#">genome.1.bt2</a>	11-Apr-2012 06:37	914M	
<a href="#">genome.2.bt2</a>	11-Apr-2012 06:38	682M	
<a href="#">genome.3.bt2</a>	11-Apr-2012 06:36	3.2K	
<a href="#">genome.4.bt2</a>	11-Apr-2012 06:38	682M	
<a href="#">genome.fa</a>	16-Mar-2012 08:46	2.9G	
<a href="#">genome.rev.1.bt2</a>	11-Apr-2012 06:37	914M	
<a href="#">genome.rev.2.bt2</a>	11-Apr-2012 06:37	682M	

Index of /NGS\_DAT/Lv2\_2/iGenome  
/Homo\_sapiens/NCBI/build37.2/Annotation  
/Archives/archive-2014-06-02-13-47-29/Genes

Name	Last modified	Size	Description
 Parent Directory	-	-	-
 genes.gtf	03-Jun-2014 07:32	102M	

ドラッグ&ドロップ

## 101ページ

☆ wgetのオプション

-r : specify recursive download  
(再帰取得)

-np : no parent

ディレクトリ階層の外部にある  
データの取得を禁止する

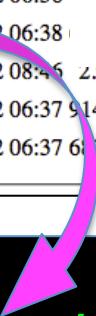
```
$ wget -r -np
```

```
$ wget http://shujunsha.com/NGS_DAT/Lv2_2/iGenome/  
Homo_sapiens/NCBI/build37.2/Annotation/Archives/  
archive-2014-06-02-13-47-29/Genes/genes.gtf
```

## **Index of /NGS\_DAT/Lv2\_2/iGenome /Homo\_sapiens/NCBI/build37.2/Sequence /Bowtie2Index**

	<a href="#">Name</a>	<a href="#">Last modified</a>	<a href="#">Size</a>	<a href="#">Description</a>
	<a href="#">Parent Directory</a>	-		
	<a href="#">genome.1.bt2</a>	11-Apr-2012 06:37	914M	
	<a href="#">genome.2.bt2</a>	11-Apr-2012 06:38	682M	
	<a href="#">genome.3.bt2</a>	11-Apr-2012 06:36		
	<a href="#">genome.4.bt2</a>	11-Apr-2012 06:38		
	<a href="#">genome.fa</a>	16-Mar-2012 08:46	2.9G	
	<a href="#">genome.rev.1.bt2</a>	11-Apr-2012 06:37	914M	
	<a href="#">genome.rev.2.bt2</a>	11-Apr-2012 06:37	682M	

ドラッグ&ドロップ



```
$ wget http://shujunsha.com/NGS_DAT/Lv2_2/iGenome/  
Homo_sapiens/NCBI/build37.2/Sequence/Bowtie2Index/  
genome.fa
```

ダウンロードにかなり時間がかかる...

# 訂正

```
$ mdfind genes.gtf -onlyin ~/expression/
```

```
/Users/Ohmomo/expression/shujunsha.com/  
NGS_DAT/Lv2_2/iGenome/Homo_sapiens/NCBI/  
build37.2/Annotation/Archives/  
archive-2014-06-02-13-47-29/Genes/genes.gtf
```

# レファレンス配列データの入手

101ページ参照



## 1. Dry解析教本専用サーバからダウンロード

([http://shujunsha.com/NGS\\_DAT/Lv2\\_2/iGenome](http://shujunsha.com/NGS_DAT/Lv2_2/iGenome))



## 2. 公開データのダウンロード

### Ensembl

(<http://www.ensembl.org/index.html>)

#### <入手するファイル>

- Fastaファイル (genome.fa)
- GTFファイル (genes.gtf)



### iGenome

([https://support.illumina.com/sequencing/sequencing\\_software/igenome.html](https://support.illumina.com/sequencing/sequencing_software/igenome.html))

Search:  for    
e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [coronary heart disease](#)

## Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

### Popular genomes



**Human**  
GRCh38.p5



**Human**  
GRCh37



**Mouse**  
GRCm38.p4



**Zebrafish**  
GRCz10

[Log in to customize this list](#)

### All genomes

-- Select a species --

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

### New!



Updated patches and gene set (GENCODE M8) for the mouse assembly (GRCm38.p4)



Ensembl supports data from external projects through [Track hubs](#)

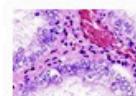
Still using Human  
GRCh37?

[Go to e!GRCh37](#)

Variant Effect Predictor



Gene expression in  
different tissues



Find SNPs and other  
variants for my gene

GIRTAATACATT  
CRTRAAAGTCT  
CTTCTAATTC  
GRAACATTTC

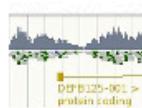
Retrieve gene sequence

GCCTGACTTCCGGTC  
GGGCTTGTGGCGGAG  
GGGCTCTGCTGGGCC  
ACGGGACAGATTGTC  
CACCTCTGAGCGCGI  
CCAGTCCAGCGTGCG

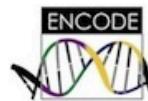
Compare genes across  
species



Use my own data in  
Ensembl



ENCODE data in Ensembl



## What's New in Ensembl Release 83 (December 2015)

- Update to Ensembl-Havana human GENCODE gene set (release 24)
- Mouse: update to Ensembl-Havana GENCODE gene set
- Chicken and pig dbSNP 145 update
- Filtering Variants by MAF
- Advanced Filtering and Counts on Variant table

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

## Latest blog posts

- 25 Jan 2016: [Sharing feature on the new mobile site \(m.ensembl.org\)](#)
- 06 Jan 2016: [What's coming in Ensembl release 84](#)
- 16 Dec 2015: [Ensembl Genomes release 30 is out!](#)

[Go to Ensembl blog](#)

## Tweets

Follow

Ensembl  
@ensembl

9h

Colib'read NGS analysis, with Ensembl splicing  
#CitedEnsembl @GigaScience @pierre350d  
@Yvan2935 buff.ly/20Dsa0l

Expand

Ensembl  
@ensembl

11h

## Using this website

## Annotation and prediction

## Data access

## API &amp; software

## About us

 Help & Documentation

## Help & Documentation

### Using this website

Our website offers lots of ways to view and interact with our genomic data - find out more!

- [Adding custom tracks](#)
- [Tutorials](#)
- [Glossary](#)
- [What's New](#)
- [FAQs](#) (Frequently Asked Questions)

[More...](#)

### Annotation & Prediction

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.



- [Gene annotation](#)
- [Variation data](#)
- [Comparative genomics](#)
- [Regulatory build](#)

[More...](#)

### Data access

All of our data is open-access and can be downloaded free of charge ([disclaimer](#)). Ways to access this data include:



- [Export](#) features or sequence directly from web pages
- Extract data from our [public database](#) using Perl scripts
- Data-mining using the [BioMart](#) tool
- [FTP download](#) of complete datasets

### API & Software

Ensembl releases all its software under an Apache-style open source [licence](#). Our products include:



- [Perl API](#) for direct data access
- [REST](#)  server for language agnostic access
- [Virtual machine](#) preloaded with API
- [Variant Effect Predictor](#) (VEP) and other command-line scripts

## FTP Download

You can download via a browser from our [FTP site](#), use a script, or even use `rsync` from the command line.

### API Code

If you do not have access to git, you can obtain our latest API code as a gzipped tarball:

[Download complete API for this release](#)

Note: the API version needs to be the same as the databases you are accessing, so please use git to obtain a previous version if querying older databases.

### Database dumps

Entire databases can be downloaded from our FTP site in a variety of formats. Please be aware that some of these files can run to many gigabytes of data.

**Looking for MySQL dumps to install databases locally?** See our [web installation instructions](#) for full details.

Each directory on [ftp.ensembl.org](ftp://ftp.ensembl.org) contains a [README](#) file, explaining the directory structure.

### Multi-species data

Database										
Comparative genomics	<a href="#">MySQL</a>	<a href="#">EMF</a>	<a href="#">MAF</a>	<a href="#">BED</a>	<a href="#">XML</a>	<a href="#">Ancestral Alleles</a>				
BioMart	<a href="#">MySQL</a>	-	-	-	-	-				
Stable ids	<a href="#">MySQL</a>	-	-	-	-	-				

### Single species data

Popular species are listed first. You can customise this list via our [home page](#).

	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig	
Y	<a href="#">Human <i>Homo sapiens</i></a>	<a href="#">FASTA</a>	<a href="#">EMBL</a>	<a href="#">GenBank</a>	<a href="#">GTF</a>	<a href="#">MySQL</a>	<a href="#">GVF</a>	<a href="#">VCF</a>	<a href="#">VEP</a>	<a href="#">Regulation</a>	<a href="#">(GFF)</a>	<a href="#">Regulation on data files</a>	<a href="#">BAM/BigWig</a>				
Y	<a href="#">Mouse <i>Mus musculus</i></a>	<a href="#">FASTA</a>	<a href="#">EMBL</a>	<a href="#">GenBank</a>	<a href="#">GTF</a>	<a href="#">MySQL</a>	<a href="#">GVF</a>	<a href="#">VCF</a>	<a href="#">VEP</a>	<a href="#">Regulation</a>	<a href="#">(GFF)</a>	<a href="#">Regulation on data files</a>	<a href="#">BAM/BigWig</a>				
Y	<a href="#">Zebrafish <i>Danio rerio</i></a>	<a href="#">FASTA</a>	<a href="#">EMBL</a>	<a href="#">GenBank</a>	<a href="#">GTF</a>	<a href="#">MySQL</a>	<a href="#">GVF</a>	<a href="#">VCF</a>	<a href="#">VEP</a>	-	-	<a href="#">BAM/BigWig</a>					
	<a href="#">Alpaca <i>Vicugna pacos</i></a>	<a href="#">FASTA</a>	<a href="#">EMBL</a>	<a href="#">GenBank</a>	<a href="#">GTF</a>	<a href="#">MySQL</a>	-	-	<a href="#">VEP</a>	-	-	<a href="#">BAM/BigWig</a>					

- `Homo_sapiens.GRCh38.dna.toplevel.fa.gz`
- `Homo_sapiens.GRCh38.83.gtf.gz`

## iGenomes

### Ready-To-Use Reference Sequences and Annotations

The iGenomes are a collection of reference sequences and annotation files for commonly analyzed organisms. The files have been downloaded from Ensembl, NCBI, or UCSC, and chromosome names have been changed to be simple and consistent with their download source. Each iGenome is available as a compressed file that contains sequences and annotation files for a single genomic build of an organism.

For more information, see the [iGenomes Overview](#) and [Change Log](#).

Species	Source	Build(s)		
<i>Arabidopsis thaliana</i>	Ensembl	TAIR10	TAIR9	
	NCBI	TAIR10	build9.1	
<i>Bacillus cereus</i> strain ATCC 10987	NCBI	2003-02-13		
<i>Bacillus subtilis</i> strain 168	Ensembl	EB2		
<i>Bos taurus</i> (Cow)	Ensembl	UMD3.1	Btau_4.0	
	NCBI	UMD_3.1.1	UMD_3.1	Btau_4.6.1
	UCSC	bosTau8	bosTau7	bosTau6
<i>Caenorhabditis elegans</i>	Ensembl	WBcel235	WBcel215	WS220
	NCBI	WS195	WS190	WS210
	UCSC	ce10	ce6	
<i>Canis familiaris</i> (Dog)	Ensembl	CanFam3.1	BROADD2	
	NCBI	build3.1	build2.1	
	UCSC	canFam3	canFam2	
<i>Danio rerio</i> (Zebrafish)	Ensembl	GRCz10	Zv9	
	NCBI	GRCz10	Zv9	
	UCSC	danRer10	danRer7	
<i>Drosophila melanogaster</i>	Ensembl	BDGP6	BDGP5	BDGP5.25
	NCBI	build5.41	build5.3	build5
	UCSC	dm6	dm3	build4.1
<i>Enterobacteriophage lambda</i>	NCBI	1993-04-28		
<i>Equus caballus</i> (Horse)	Ensembl	EquCab2		
	NCBI	EquCab2.0		
	UCSC	equCab2		
<i>Escherichia coli</i> strain K12, DH10B	Ensembl	EB1		
	NCBI	2008-03-17		
<i>Escherichia coli</i> strain K12, MG1655	NCBI	2001-10-15		
<i>Gallus gallus</i> (Chicken)	Ensembl	Galgal4	WASHUC2	
	NCBI	build3.1	build2.1	
	UCSC	galGal4	galGal3	
<i>Glycine max</i>	Ensembl	Gm01		
<i>Homo sapiens</i>	Ensembl	GRCh37		
	NCBI	GRCh38	build37.2	build37.1
	UCSC	hg38	hg19	hg18

# 圧縮ファイルから必要なファイルのみを解凍する方法

103ページ参照

# 解析データの入手

1. fastq.gzファイルをダウンロード
2. sraファイルをダウンロード  
→ fastqファイルへの変換が必要

```
$ cd ~/expression/
```

[http://www.ebi.ac.uk/arrayexpress/  
experiments/E-MTAB-1086/](http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1086/)




Search

Advanced

ArrayExpress &gt; Browse &gt; E-MTAB-1086

## E-MTAB-1086 - ChIP-seq and RNA-seq of coding RNA of the progression of human embryonic stem cells to beta cells to characterize the epigenetic programs that underlie pancreas differentiation

Status *Released on 8 February 2013, last updated on 3 May 2014*

Organism *Homo sapiens*

Samples (7) [Click for detailed sample information and links to data](#)

Protocols (14) [Click for detailed protocol information](#)

Description To characterize the epigenetic programs that underlie pancreas differentiation, we have generated genome-scale maps of H3K4me and H3K27me3 patterns by ChIP-seq and determined expression profiles by RNA-seq from undifferentiated human ESCs, four intermediate differentiated stages (definitive endoderm, primitive gut tube, posterior foregut, and pancreatic endoderm), and in vitro-differentiated polyhormonal cells. Antibodies against CD142 and CD200 were used to select for targeted pancreatic and endocrine populations at the end of the culture. Cells at the end of culture were implanted into mice for further differentiation into mature insulin-producing beta-cells and compared to sorted polyhormonal cells by RNA-seq and ChIP-seq analysis. For the experimental factor values the time points are along a differentiation protocol from stem cells to tissue. Values in the CD antibody column refer to antibodies used for cell sorting. In this column 'not applicable' refers to samples that were not sorted. 'none' refers to samples that were sorted, but did not bind to either of the two antibodies, CD200 and CD142. Values in the histone antibody column are about the ChIP process.

Experiment types ChIP-seq, RNA-seq of coding RNA, binding site identification, cell type comparison, co-expression, development or differentiation, is expressed

Contacts Maike Sander <[masander@ucsd.edu](mailto:masander@ucsd.edu)>, Jonathan Schug <[jschug@mail.med.upenn.edu](mailto:jschug@mail.med.upenn.edu)>, Ruiyu Xie <[rxi@ucsd.edu](mailto:rxi@ucsd.edu)>

Citation [Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells](#). Xie R, Everett LJ, Lim HW, Patel NA, Schug J, Kroon E, Kelly OG, Wang A, D'Amour KA, Robins AJ, Won KJ, Kaestner KH, Sander M. , [Europe PMC 23318056](#)

MINSEQE \* - \* \* \*  
Exp. design Protocols Variables Processed Seq. reads

Files Investigation description  [E-MTAB-1086.idf.txt](#)  
 Sample and data relationship  [E-MTAB-1086.sdrf.txt](#)  
 Processed data (1)  [E-MTAB-1086.processed.1.zip](#)  
[Click to browse all available files](#)

Links [ENA - ERP002045](#)  
[Send E-MTAB-1086 data to GENOME SPACE](#)



ArrayExpress &gt; Browse &gt; E-MTAB-1086 &gt; Samples and Data\*

E-MTAB-1086 - ChIP-seq and RNA-seq of coding RNA of the progression of human embryonic stem cells to beta cells to characterize the epigenetic programs that underlie pancreas differentiation

Source Name	Sample Attributes						CD a
	organism	developmental stage	cell line	cell type	TIMEPOINT		
Cohort (E2147)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
Cohort (E2147)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
Cohort (E2147)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
Cohort (E2147)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
Cohort (E2147)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
Cohort (E2182)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
Cohort (E2182)	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 13	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 2	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 2	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 2	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 0	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 0	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 5	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 7	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 0	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 5	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 7	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 5	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 10	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 10	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 7	n/a	
DP1	Homo sapiens	embryo	Cyt49	embryonic stem cell	day 10	n/a	

Links to Data			Links to Data		
ENA	FASTQ	Processed	ENA	FASTQ	Processed
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>	<a href="#"></a>
<a href="#"></a>	<a href="#">&lt;/a</a>				

DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 0	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 2	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 5	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 7	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 10	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 0	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 2	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 5	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 7	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell	day 10	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell	day 0	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell	day 2	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell	day 5	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell	day 7	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell	day 10	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD142-PE, BD 550312, lot#: 78463	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD142-PE, BD 550312, lot#: 78463	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K4me3 Millipore 04-745	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K27me3 Millipore 07-441	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD142-PE, BD 550312, lot#: 78463	H3K4me3 Millipore 04-745	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD142-PE, BD 550312, lot#: 78463	H3K27me3 Millipore 07-441	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	CD142-PE, BD 550312, lot#: 78463	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	none	input	<a href="#">🔗</a>	<a href="#">⬇️</a>	-
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell	day 13	none	input	<a href="#">🔗</a>	<a href="#">⬇️</a>	-
HI40	Homo sapiens adult		islet cell	n/a	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>
HI41	Homo sapiens adult		islet cell	n/a	n/a	n/a (rna-seq)	<a href="#">🔗</a>	<a href="#">⬇️</a>	<a href="#">⬇️</a>

[Download Samples and Data table in Tab-delimited format](#)

Examples: BN000065, histone

[Search](#)  
[Advanced](#)  
[Sequence](#)
[Home](#) | [Search & Browse](#) | [Submit & Update](#) | [About ENA](#) | [Support](#)

- ⓘ SVA will be unavailable Thursday 18th February  
 o Please subscribe to ena-announce mailing list here:[listserver.ebi.ac.uk/mailman/listinfo/ena-announce](mailto:listserver.ebi.ac.uk/mailman/listinfo/ena-announce) to receive alerts about ENA services.

## Run: ERR266335

Illumina HiSeq 2000 sequencing; Transcriptional and epigenetic profiling of the progression of hESCs to beta cells

 View: [XML](#)
[Send Feedback](#)
[Download: XML](#)

Submitting Centre	Run Date	Platform	Model	Read Count	Base Count
PEDIATRIC DIABETES RESEARCH CENTER, UCSD STEM CELL PROGRAM		ILLUMINA	Illumina HiSeq 2000	64,533,918	6,453,391,800
Library Layout	Library Strategy	Library Source	Library Selection	Library Name	
SINGLE	RNA-Seq	TRANSCRIPTOMIC	cDNA	2010	
Broker Name	<a href="#">ArrayExpress</a>				

[Navigation](#) [Read Files](#)

This table contains the files for run ERR266335

[Download files](#)

 Download:  -  of 1 results in [TEXT](#)
[Select columns](#)

Showing results 1 - 1 of 1 results

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)	Submitted files (ftp)	Submitted files (galaxy)	CRAM Index files (ftp)	CRAM Index files (galaxy)
PRJEB1195	ERP002045	SAMEA1712776	ERS197530	ERX182652	ERR266335	9606	Homo sapiens	Illumina HiSeq 2000	SINGLE	File 1	File 1	Fastq file 1	Fastq file 1		

[Prev](#) [Next](#) Go to page:  Go

# 解析データの入手

DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 0	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 2	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 5	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 7	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 10	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 0	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 2	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 5	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 7	n/a	n/a (rna-seq)			
DP1	Homo sapiens embryo	Cyt49	embryonic stem cell day 10	n/a	n/a (rna-seq)			
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell day 0	n/a	n/a (rna-seq)			
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell day 2	n/a	n/a (rna-seq)			
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell day 5	n/a	n/a (rna-seq)			
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell day 7	n/a	n/a (rna-seq)			
DP3	Homo sapiens embryo	Cyt49	embryonic stem cell day 10	n/a	n/a (rna-seq)			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	n/a (rna-seq)			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	n/a (rna-seq)			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD142-PE, BD 550312, lot#: 78463	n/a (rna-seq)			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD142-PE, BD 550312, lot#: 78463	n/a (rna-seq)			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K4me3 Millipore 04-745			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K27me3 Millipore 07-449			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K4me3 Millipore 04-745			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K27me3 Millipore 07-449			
DP3, 4, 5	Homo sapiens embryo	Cyt49	embryonic stem cell day 13	CD200-APC, eBioscience 17-9200, lot# E12743-101	H3K4me3 Millipore 04-745			
HI40	Homo sapiens adult	islet			n/a (rna-seq)			
HI41	Homo sapiens adult	islet			n/a (rna-seq)			

合計約43GBもあるので時間がかかる

その後ろに  
アップ

# 解析データの確認

findコマンドを使って  
解析データを確認しよう。

105ページ参照

# fastq.gzの解凍

gunzipコマンドを使って  
解析データを解凍しよう。

105ページ参照

# 解析データの入手

1. fastq.gzファイルをダウンロード
2. sraファイルをダウンロード  
→ fastqファイルへの変換が必要

sra toolkitをインストールしよう

106ページ参照

# fastqファイルの中身を確認

```
$ head ERR266335.fastq
```



- ① : 機器ID、レーン番号、タイル番号、タイル内のクラスターのX座標、タイル内のクラスターのY座標、インデックス番号、ペアの番号
  - ② : 塩基配列
  - ③ : + (配列のIDを記載することもある)
  - ④ : ②の配列のクオリティスコア

# FastQCのインストール...の前に

fastqファイルのクオリティをチェックするためのソフト。Javaで動くため、Javaをダウンロードしておく必要がある。



The screenshot shows the OTN Java SE Downloads page. The top navigation bar includes links for Sign In/Register, Help, Country, Communities, I am a..., I want to..., Search, Products, Solutions, Downloads, Store, Support, Training, Partners, About, and OTN. The main menu also lists Java SE, Java EE, Java ME, Java Card, Java TV, New to Java, and Community. On the left, there's a sidebar with links for Java SE, Java EE, Java ME, Java SE Support, Java SE Advanced & Suite, Java Embedded, Java DB, Web Tier, Java Card, Java TV, New to Java, and Community. The right sidebar contains sections for Java SDKs and Tools (Java SE, Java EE and Glassfish, Java ME, Java Card, NetBeans IDE, Java Mission Control), Java Resources (Java APIs, Technical Articles, Demos and Videos, Forums, Java Magazine, Java.net, Developer Training, Tutorials, Java.com), and Java Platform, Standard Edition (Java SE 8u73 / 8u74, Java Platform, Standard Edition). The central content area features two download cards: 'Java Platform (JDK) 8u73 / 8u74' (with a pink box around the 'DOWNLOAD' button) and 'NetBeans with JDK 8'. Below these are detailed descriptions for Java SE 8u73 / 8u74 and Java Platform, Standard Edition, along with links for Installation Instructions, Release Notes, and Oracle License.

## Java SE Development Kit 8u73

You must accept the Oracle Binary Code License Agreement for Java SE to download this software.

Accept License Agreement  Decline License Agreement

Product / File Description	File Size	Download
Linux ARM v6/v7 Hard Float ABI	77.73 MB	<a href="#">jdk-8u73-linux-arm32-vfp-hflt.tar.gz</a>
Linux ARM v6/v7 Hard Float ABI	74.68 MB	<a href="#">jdk-8u73-linux-arm64-vfp-hflt.tar.gz</a>
Linux x86	154.75 MB	<a href="#">jdk-8u73-linux-i586.rpm</a>
Linux x86	174.91 MB	<a href="#">jdk-8u73-linux-i586.tar.gz</a>
Linux x64	152.73 MB	<a href="#">jdk-8u73-linux-x64.rpm</a>
Linux x64	172.91 MB	<a href="#">jdk-8u73-linux-x64.tar.gz</a>
Mac OS X x64	227.25 MB	<a href="#">jdk-8u73-macosx-x64.dmg</a>
Solaris SPARC 64-bit (SVR4 package)	139.7 MB	<a href="#">jdk-8u73-solaris-sparcv9.tar.Z</a>
Solaris SPARC 64-bit	99.08 MB	<a href="#">jdk-8u73-solaris-sparcv9.tar.gz</a>
Solaris x64 (SVR4 package)	140.36 MB	<a href="#">jdk-8u73-solaris-x64.tar.Z</a>
Solaris x64	96.78 MB	<a href="#">jdk-8u73-solaris-x64.tar.gz</a>
Windows x86	181.5 MB	<a href="#">jdk-8u73-windows-i586.exe</a>
Windows x64	186.84 MB	<a href="#">jdk-8u73-windows-x64.exe</a>

## Java SE Development Kit 8u74

You must accept the Oracle Binary Code License Agreement for Java SE to download this software.

Accept License Agreement  Decline License Agreement

Product / File Description	File Size	Download
Linux x86	154.74 MB	<a href="#">jdk-8u74-linux-i586.rpm</a>
Linux x86	174.92 MB	<a href="#">jdk-8u74-linux-i586.tar.gz</a>
Linux x64	152.74 MB	<a href="#">jdk-8u74-linux-x64.rpm</a>
Linux x64	172.9 MB	<a href="#">jdk-8u74-linux-x64.tar.gz</a>
Mac OS X x64	227.27 MB	<a href="#">jdk-8u74-macosx-x64.dmg</a>
Solaris SPARC 64-bit (SVR4 package)	139.72 MB	<a href="#">jdk-8u74-solaris-sparcv9.tar.Z</a>
Solaris SPARC 64-bit	99.09 MB	<a href="#">jdk-8u74-solaris-sparcv9.tar.gz</a>
Solaris x64 (SVR4 package)	140.02 MB	<a href="#">jdk-8u74-solaris-x64.tar.Z</a>
Solaris x64	96.19 MB	<a href="#">jdk-8u74-solaris-x64.tar.gz</a>
Windows x86	182.01 MB	<a href="#">jdk-8u74-windows-i586.exe</a>
Windows x64	187.31 MB	<a href="#">jdk-8u74-windows-x64.exe</a>

クリックして  
ダウンロードする





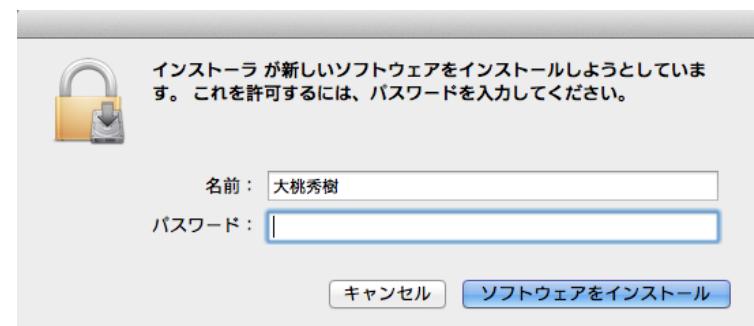
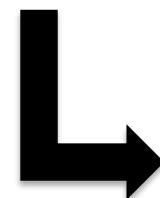
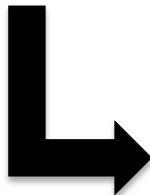
JDK 8 Update 74

## Java Development Kit

Double-click on icon to install



JDK 8 Update 74.pkg



109ページ参照

## FastQCのインストールしよう

## FastQC結果を入れるディレクトリを作成しよう

## FastQCのオプションを確認しよう

```
$ fastqc -h
```

## SYNOPSIS

```
fastqc seqfile1 seqfile2 .. seqfileN
fastqc [-o output dir] [--(no)extract] [-f fastq|bam|sam]
[-c contaminant file] seqfile1 .. seqfileN
```

## DESCRIPTION

FastQC reads a set of sequence files and produces from each one a quality control report consisting of a number of different modules, each one of which will help to identify a different potential type of problem in your data.

If no files to process are specified on the command line then the program will start as an interactive graphical application. If files are provided on the command line then the program will run with no user interaction required. In this mode it is suitable for inclusion into a standardised analysis pipeline.

The options for the program are as follows:

-h --help	Print this help file and exit
-v --version	Print the version of the program and exit
-o --outdir	Create all output files in the specified output directory. Please note that this directory must exist as the program will not create it. If this option is not set then the output file for each sequence file is created in the same directory as the sequence file which was processed.
--casava	Files come from raw casava output. Files in the same sample group (differing only by the group number) will be analysed as a set rather than individually. Sequences with the filter flag set in the header will be excluded from the analysis. Files must have the same names given to them by casava (including being gzipped and ending with .gz) otherwise they won't be grouped together correctly.
--nofilter	If running with --casava then don't remove read flagged by casava as poor quality when performing the QC analysis.
--extract	If set then the zipped output file will be uncompressed in the same directory after it has been created. By default this option will be set if fastqc is run in non-interactive mode.
-j --java	Provides the full path to the java binary you want to use to launch fastqc. If not supplied then java is assumed to be in your path.
--noextract	Do not uncompress the output file after creating it. You should set this option if you do not wish to uncompress the output when running in non-interactive mode.
--nogroup	Disable grouping of bases for reads >50bp. All reports will show data for every base in the read. WARNING: Using this option will cause fastqc to crash and burn if you use it on really long reads, and your plots may end up a ridiculous size. You have been warned!

## ☆ fastqcのオプション

## --nogroup :

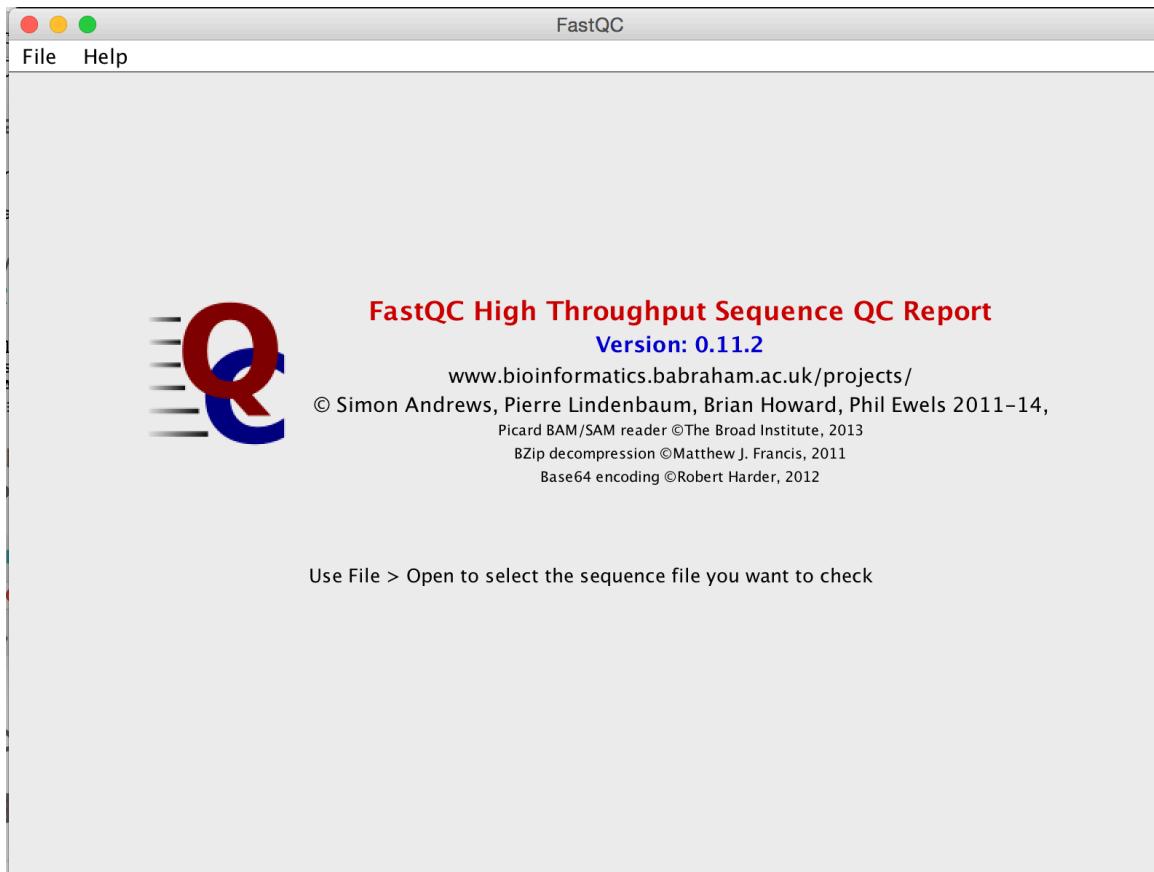
シークエンスリードが長くなると、3'末端の解析結果が束ねられてしまい、シークエンス結果が曖昧になってしまふのを防ぐコマンド

# FastQCの実行

109ページ参照

## 例) ERR266335.fasta

※ターミナルに fastqc と入力・実行すると  
GUI画面が立ち上がる



# FastQC結果

- FastQCディレクトリの中を確認

```
$ cd FastQC
```

```
$ ls
```

```
$ open -a firefox ERR266335_fastqc.html
```

Safari

Google\ Chrome

# FastQC結果の項目

- Basic Statistics : 結果のサマリー
- Per base sequence quality : 各位置におけるクオリティ
- Per tile sequence quality : タイル単位でのクオリティ
- Per sequence quality scores : 各クオリティスコアのリードの分布
- Per base sequence content : 各位置における塩基の出現頻度
- Per sequence GC content : 配列のGC 含量
- Per base N content : 各位置に現れる N の割合
- Sequence Length Distribution : リード長の分布
- Sequence Duplication Levels : Duplicated readsの割合
- Overrepresented sequences : 重複数の多いリード
- Adapter Content : アダプター含量
- Kmer Content : リードから観測される K-mer

# fastx\_toolkit

111ページ参照

クオリティの良いリード（データ）を選別するため、アダプター配列や短いリードなどを除去する

fastx\_toolkitをインストール  
しよう

fastq\_toolkitを実行しよう

# 遺伝子発現解析(3)

- RNAスプライシングアライメント  
(TopHat)
- マッピングデータのフォーマットの理解

# TopHatでマッピング、の前に

expressionディレクトリへ移動する

```
$ cd  
$ cd expression
```

```
$ find * | grep trim
```

```
$ find * | grep gtf
```

```
$ find * | grep genome.fa
```

訂正

```
$ find * | grep gtf  
shujunsha.com/NGS_DAT/Lv2_2/iGenome/Homo_sapiens/  
NCBI/build37.2/Annotation/Archives/  
archive-2014-06-02-13-47-29/Genes/genes.gtf
```

```
$ find * | grep genome.fa  
shujunsha.com/NGS_DAT/Lv2_2/iGenome/Homo_sapiens/  
NCBI/build37.2/Sequence/Bowtie2Index/genome.fa
```

# TopHat結果を入れるディレクトリを作成しよう

```
$ mkdir tophat_results
```

TopHatを仕掛けよう 112ページ参照

はたしてTopHat結果は...

# TopHat結果を確認しよう

- accepted\_hits.bam ← cufflinksで使用する  
のでファイル名を変えよう
- align\_summary.txt
- deletions.bed
- insertions.bed
- junctions.bed
- logs (ディレクトリ)
- prep\_reads.info
- unmapped.bam

133ページ